big data and cognitive computing

Knowledge Modelling and Learning through Cognitive Networks

Edited by
Massimo Stella and Yoed N. Kenett

Printed Edition of the Special Issue Published in *BDCC*

www.mdpi.com/journal/BDCC

MDPI

Knowledge Modelling and Learning through Cognitive Networks

Knowledge Modelling and Learning through Cognitive Networks

Editors

Massimo Stella
Yoed N. Kenett

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors
Massimo Stella
CogNosco Lab, Department of Computer Science, University of Exeter,
Exeter EX4 4PY, UK
Yoed N. Kenett
Faculty of Industrial Engineering and Management, Technion—Israel institute of Technology,
Haifa 3200003, Israel

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Actuators* (ISSN 2076-0825) (available at: https://www.mdpi.com/journal/BDCC/special_issues/knowledge_modelling).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-4345-1 (Hbk)
ISBN 978-3-0365-4346-8 (PDF)

Contents

About the Editors . . . . . **vii**

Massimo Stella and Yoed N. Kenett
Knowledge Modelling and Learning through Cognitive Networks
Reprinted from: *Big Data Cogn. Comput.* **2022**, *6*, 53, doi:10.3390/bdcc6020053 . . . . . **1**

Hossein Hassani,Christina Beneki,Stephan Unger,Maedeh Taj MazinaniandMohammad Reza Yeganegi
Text Mining in Big Data Analytics
Reprinted from: *Big Data Cogn. Comput.* **2020**, *4*, 1, doi:10.3390/bdcc4010001 . . . . . **5**

Birgitta Dresp-Langley
Seven Properties of Self-Organization in the Human Brain
Reprinted from: *Big Data Cogn. Comput.* **2020**, *4*, 10, doi:10.3390/bdcc4020010 . . . . . **39**

Massimo Stella, Valerio Restocchi and Simon De Deyne
#*lockdown*: Network-Enhanced Emotional Profiling in the Time of COVID-19
Reprinted from: *Big Data Cogn. Comput.* **2020**, *4*, 14, doi:10.3390/bdcc4020014 . . . . . **57**

Toni Pano and Rasha Kashef
A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19
Reprinted from: *Big Data Cogn. Comput.* **2020**, *4*, 33, doi:10.3390/bdcc4040033 . . . . . **81**

Karin Nilsson, Lisa Palmqvist, Magnus Ivarsson, Anna Levén, Henrik Danielsson, Marie Annell, Daniel Schöld and Michaela Socher
Structural Differences of the Semantic Network in Adolescents with Intellectual Disability
Reprinted from: *Big Data Cogn. Comput.* **2021**, *5*, 25, doi:10.3390/bdcc5020025 . . . . . **99**

Bhargav Prakash, Gautam Kumar Baboo and Veeky Baths
A Novel Approach to Learning Models on EEG Data Using Graph Theory Features—A Comparative Study
Reprinted from: *Big Data Cogn. Comput.* **2021**, *5*, 39, doi:10.3390/bdcc5030039 . . . . . **111**

Michael S. Vitevitch, Leo Niehorster-Cook and Sasha Niehorster-Cook
Exploring How Phonotactic Knowledge Can Be Represented in Cognitive Networks
Reprinted from: *Big Data Cogn. Comput.* **2021**, *5*, 47, doi:10.3390/bdcc5040047 . . . . . **127**

Yusuf Sermet and Ibrahim Demir
A Semantic Web Framework for Automated Smart Assistants: A Case Study for Public Health
Reprinted from: *Big Data Cogn. Comput.* **2021**, *5*, 57, doi:10.3390/bdcc5040057 . . . . . **145**

Asra Fatima, Ying Li, Thomas Hills and Massimo Stella
DASentimental: Detecting Depression, Anxiety, and Stress in Texts via Emotional Recall, Cognitive Networks, and Machine Learning
Reprinted from: *Big Data Cogn. Comput.* **2021**, *5*, 77, doi:10.3390/bdcc5040077 . . . . . **165**

Alexander Sboev, Anton Selivanov, Ivan Moloshnikov, Roman Rybka, Artem Gryaznov, Sanna Sboeva and Gleb Rylkov
Extraction of the Relations among Significant Pharmacological Entities in Russian-Language Reviews of Internet Users on Medications
Reprinted from: *Big Data Cogn. Comput.* **2022**, *6*, 10, doi:10.3390/bdcc6010010 . . . . . **183**

Arjun M. Kumar, Jasmine Y. Q. Goh, Tiffany H. H. Tan and Cynthia S. Q. Siew
Gender Stereotypes in Hollywood Movies and Their Evolution over Time: Insights from Network Analysis
Reprinted from: *Big Data Cogn. Comput.* **2022**, *6*, 50, doi:10.3390/bdcc6020050 . . . . . **199**

About the Editors

Massimo Stella

Massimo Stella, PhD is head of CogNosco Lab at the Dept. of Computer Science, University of Exeter. His research revolves around cognitive data science, developing cutting-edge artificial intelligence methods grounded in cognitive science and complex networks. His lab works on multiple applications of natural language processing in clinical settings, social media analysis, and innovation in education.

Yoed N. Kenett

Yoed N. Kenett PhD leads the Cognitive Complexity Lab at the faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology. His research focuses on the complexity of high-level cognition, such as knowledge, creativity, associative thought, question asking, and memory search in typical and atypical populations. To study these issues, he applies network science methodologies at the cognitive and neural levels, converging computational and empirical research.

big data and cognitive computing

MDPI

Editorial

Knowledge Modelling and Learning through Cognitive Networks

Massimo Stella [1,*] and Yoed N. Kenett [2,*]

[1] CogNosco Lab, Department of Computer Science, University of Exeter, Exeter EX4 4PY, UK
[2] Faculty of Industrial Engineering and Management, Technion-Israel Institute of Technology, Haifa 3200003, Israel
* Correspondence: m.stella@exeter.ac.uk (M.S.); yoedk@technion.ac.il (Y.N.K.)

Knowledge modelling is a growing field at the fringe of computer science, psychology and network science [1,2]. This research area aims to build models of knowledge that can provide interpretable insights starting from data, its associations, commonalities, recurrent patterns and correlations. Historically, artificial intelligence (AI) contributed vastly to the field through models like artificial neural networks, e.g., recurrent neural networks or deep learning, as methods able to extract knowledge and learn from data, cf. [1]. Recent advancements from fields like network and data science supported the creation of novel approaches to knowledge modelling, inspired by theoretical frameworks of cognition and language processing: cognitive networks are mental representations of knowledge where nodes represent concepts and links indicate conceptual associations, e.g., concepts sounding similarly or being related according to a given semantic definition, cf. [3,4].

Despite being both referred to as "networks", artificial neural networks (ANNs) and cognitive networks (CNs) remain two frameworks that work well in synergy while remaining distinct. On the one hand, ANNs encapsulate in their network structure latent correlations in the data, making it difficult to identify what nodes and their interconnections represent [5]. On the other hand, CNs form one-to-one mappings of knowledge units, e.g., nodes represent specific concepts and links map specific types of conceptual associations [4]. Whereas CNs are evidently more interpretable and can be tuned to map specific aspects of human associative knowledge (e.g., semantic memory structure and its influence over cognitive traits [6]), CNs also lack the same generalisability and aptitude to learn from data that ANNs possess, also thanks to training and fine-tuning [5]. Another important difference is that ANNs focus on prediction by updating weights between layers while CNs focus on the representation of the complexity of systems via graphs [3,6].

The synergy of these two approaches can open new ways for modelling knowledge and learning in interpretable ways, able to account also for unseen data [3,6]. For instance, the structure of CNs can produce novel features that can then power artificial intelligence techniques inspired by human knowledge, with relevant advancements for natural language processing, automatic assessments of personality traits or other phenomena like emotional distress, as highlighted in all the papers published within this Special Issue (SI).

The current SI reports on recent developments in applying CNs and ANNs for achieving intelligent systems and data insights. This SI represents a multidisciplinary collection of 11 contributions using either CNs, ANNs or novel combinations and mainly organised along the lines of: (i) text processing and social media analysis, (ii) artificial intelligence for natural language processing and (iii) brain science and cognitive psychology.

Hassani and colleagues [C1] reviewed text mining techniques for understanding features of texts in large volumes and with the assistance of quantitative AI techniques. The authors also reviewed cutting-edge methods for understanding text sentiment (valence for psychologists), i.e., pleasantness/displeasure as expressed in language. The review critically covered the many advances in the field and underlined the need for

Citation: Stella, M.; Kenett, Y.N. Knowledge Modelling and Learning through Cognitive Networks. *Big Data Cogn. Comput.* **2022**, *6*, 53. https://doi.org/10.3390/bdcc6020053

Received: 7 May 2022
Accepted: 11 May 2022
Published: 13 May 2022

novel cognitively-inspired methods, building a bridge between words in texts and ideas in the mind.

Stella and colleagues [C2] introduced network-based methods for identifying not only sentiment but also emotions in social media data. Focusing on the Italian twittersphere in the aftermath of the first COVID-19 lockdown, the authors reconstructed online stances of COVID-19 related hashtags and their emotional profiles. Emotional states as complex as trust, fear and anger were found to surround the same hashtag in different ways, according to contextual knowledge that was modelled as a cognitive network.

Pano and Kashef [C3] worked on COVID-19 tweets but under the perspective of monitoring conversations explicitly related to bitcoin. The authors tested 13 strategies for correlating textual data with bitcoin prices and identified a list of methodological working assumptions and limitations affecting predictions, showcasing a link between social discourse and price fluctuations but only in small time spans.

Prakash and colleagues [C4] used human-centred machine learning to predict the efficacy of treatment out of CNs built from brain data. The authors showed how recurrent neural networks were able to learn network features and achieve an accuracy of almost 78% in correctly classifying individuals according to their self-perceived efficacy of treatment. These findings open the way to novel ways for measuring psychological constructs from brain data, contributing to bridging the brain and mind aspects of human cognition.

Sermet and Demir [C5] outlined how cognitive, textual and social data might be combined in the AI pipeline of smart assistants, i.e., AI extracting insights from input data, predicting trends and managing conversations in natural language. The authors underlined how their cognitive computing approach enabled reusability and reliability and also discussed the relevance of their smart assistant in managing COVID-19 health data.

Sboev and colleagues [C6] introduced a context-dependent framework enforcing contextual semantic features of concepts in texts in an interpretable way and in synergy with pre-existing transformer networks. The authors' approach enables natural language processing where explicit features of language and context are both accessible to experimenters, improving model interpretability and also performance. Deploying their architecture in medical reports, the authors report on the importance of contextual features over accurate predictions.

Fatima and colleagues [C7] used network features and recurrent neural networks to predict psychological constructs, i.e., depression, anxiety and stress. The authors used emotional recalls and psychometric data to train an AI in spotting depression, anxiety and stress levels out of word combinations. Their cognitive embedding assessed word centrality and semantic distance in a network representation of associative knowledge between 36k English words. The authors validated the AI on a set of suicide notes through the circumplex model of affect.

Nilsson and colleagues [C8] used CNs to model the mental lexicon of children with typical development and adolescents with intellectual disabilities. The authors found that adolescents with intellectual disabilities produced less modular, more clustered and less spread apart layouts of conceptual associations, clustering concepts more than children with typical development. The authors also discussed the interpretation of these differences and the potential role played by context and education.

Dresp-Langley [C9] used network features—related to connectivity, resilience and information processing—as explorative dimensions for self-organisation, i.e., the ability for a system to evolve dynamically towards a working conformation. The author showed how brain networks evolve towards self-organisation while minimising system complexity and enhancing its resilience and adaptiveness, with implications also for cognitive computing.

Vitevitch and colleagues [C10] used numerical simulations to bridge together CN structure and language processing. The authors explored three representations of human memory based on different phonological similarities between concepts. Simulations showed how activation spreading across network links could account for many effects observed in empirical experiments about phonotactic knowledge and affecting spoken

word recognition. Their work underlines how cognitive networks can effectively model and test processes relative to language understanding and use.

Siew and colleagues [C11] adopted CNs to investigate the presence of stereotypical socio-cognitive representations of gender roles within Western movies from 1940 to 2019. The authors used word co-occurrences in movie synopses to capture syntactic relationships and semantic frames, blending natural language processing and cognitive network science methods. Their analysis identified the prevalence of stereotypical representations of female characters, being more entrenched in family and romance jargon than male counterparts. This approach opens new ways to quantify gender stereotypes as represented in cultural products.

Overall, our SI demonstrates the strengths and great potential of converging network science, data science, natural language processing, machine learning, and artificial intelligence to study knowledge representation and phenomena. Human knowledge is a complex system that traditionally was only examined indirectly. The expedited advancement in computational and analytical methodologies is rapidly advancing our understanding of its complexity. Our SI is what we hope is just one step forward in such a direction, a direction that harnesses state-of-the-art computational tools in the quest to better understand human knowledge.

List of Contributors:

C1. Hassani, H.; Beneki, C.; Unger, S.; Mazinani, M.T.; Yeganegi, M.R. Text Mining in Big Data Analytics. *Big Data Cogn. Comput.* **2020**, *4*, 1.
C2. Stella, M.; Restocchi, V.; De Deyne, S. #*lockdown*: Network-Enhanced Emotional Profiling in the Time of COVID-19. *Big Data Cogn. Comput.* **2020**, *4*, 14.
C3. Pano, T.; Kashef, R.A. Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data Cogn. Comput.* **2020**, *4*, 33.
C4. Prakash, B.; Baboo, G.K.; Baths, V. A Novel Approach to Learning Models on EEG Data Using Graph Theory Features—A Comparative Study. *Big Data Cogn. Comput.* **2021**, *5*, 39.
C5. Sermet, Y.; Demir, I. A Semantic Web Framework for Automated Smart Assistants: A Case Study for Public Health. *Big Data Cogn. Comput.* **2021**, *5*, 57.
C6. Sboev, A.; Selivanov, A.; Moloshnikov, I.; Rybka, R.; Gryaznov, A.; Sboeva, S.; Rylkov, G. Extraction of the Relations among Significant Pharmacological Entities in Russian-Language Reviews of Internet Users on Medications. *Big Data Cogn. Comput.* **2022**, *6*, 10.
C7. Fatima, A.; Li, Y.; Hills, T.T.; Stella, M. DASentimental: Detecting Depression, Anxiety, and Stress in Texts via Emotional Recall, Cognitive Networks, and Machine Learning. *Big Data Cogn. Comput*. **2021**, *5*, 77.
C8. Nilsson, K.; Palmqvist, L.; Ivarsson, M.; Levén, A.; Danielsson, H.; Annell, M.; Schöld, D.; Socher, M. Structural Differences of the Semantic Network in Adolescents with Intellectual Disability. *Big Data Cogn. Comput.* **2021**, *5*, 25.
C9. Dresp-Langley, B. Seven Properties of Self-Organization in the Human Brain. *Big Data Cogn. Comput*. **2020**, *4*, 10.
C10. Vitevitch, M.S.; Niehorster-Cook, L.; Niehorster-Cook, S. Exploring How Phonotactic Knowledge Can Be Represented in Cognitive Networks. *Big Data Cogn. Comput.* 2021, 5, 47.
C11. Kumar, A.M., Goh, J.Y.Q., Tan, T.H.H., Siew, C.S.Q. Gender Stereotypes in Western Movies and Their Evolution Over Time: Insights from Network Analysis. *Big Data Cogn. Comput.* **2022**, *6*, 50

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fensel, D. *Ontologies*; Springer: Berlin, Germany, 2001; pp. 11–18.
2. Stella, M. Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media. *PeerJ Comput. Sci.* **2020**, *6*, e295. [CrossRef]
3. Hills, T.T.; Kenett, Y.N. Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. *Top. Cogn. Sci.* **2022**, *14*, 189–208. [CrossRef]
4. Siew, C.S.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, *2109*, 2108423. [CrossRef]

5. Aggarwal, C.C. *Neural Networks and Deep Learning*; Springer: San Francisco, CA, USA, 2018; Volume 10, p. 978-3.
6. Stella, M.; Kenett, Y.N. Viability in multiplex lexical networks and machine learning characterizes human creativity. *Big Data Cogn. Comput.* **2019**, *3*, 45. [CrossRef]

big data and cognitive computing

MDPI

Article

Text Mining in Big Data Analytics

Hossein Hassani [1,*], Christina Beneki [2], Stephan Unger [3] and Maedeh Taj Mazinani [4] and Mohammad Reza Yeganegi [5]

1 Research Institute of Energy Management and Planning, University of Tehran, Tehran 1417466191, Iran
2 Department of Tourism, Faculty of Economic Sciences, Ionian University, Kalypso Building, 4 P. Vraila Armeni, 49100 Corfu, Greece; benekic@ionio.gr
3 Department of Economics and Business, Saint Anselm College, 100 Saint Anselm Drive, Manchester, NH 03103, USA; sunger@anselm.edu
4 Department of Management, University of Tehran, Tehran 1417466191, Iran; maedetaj@ut.ac.ir
5 Department of Accounting, Islamic Azad University, Central Tehran Branch, Tehran 1955847781, Iran; m.yeganegi@iauctb.ac.ir
* Correspondence: hassani.stat@gmail.com

Received: 18 November 2019; Accepted: 11 January 2020; Published: 16 January 2020

Abstract: Text mining in big data analytics is emerging as a powerful tool for harnessing the power of unstructured textual data by analyzing it to extract new knowledge and to identify significant patterns and correlations hidden in the data. This study seeks to determine the state of text mining research by examining the developments within published literature over past years and provide valuable insights for practitioners and researchers on the predominant trends, methods, and applications of text mining research. In accordance with this, more than 200 academic journal articles on the subject are included and discussed in this review; the state-of-the-art text mining approaches and techniques used for analyzing transcripts and speeches, meeting transcripts, and academic journal articles, as well as websites, emails, blogs, and social media platforms, across a broad range of application areas are also investigated. Additionally, the benefits and challenges related to text mining are also briefly outlined.

Keywords: text mining; big data; analytics; review

1. Introduction

In recent years, we have witnessed an increase in the quantities of available digital textual data, generating new insights and thereby opening up opportunities for research along new channels. In this rapidly evolving field of big data analytic techniques, text mining has gained significant attention across a broad range of applications. In both academia and industry, there has been a shift towards research projects and more complex research questions that mandate more than the simple retrieval of data. Due to the increasing importance of artificial intelligence and its implementation on digital platforms, the application of parallel processing, deep learning, and pattern recognition to textual information is crucial. All types of business models, market research, marketing plans, political campaigns, or strategic decision-making are facing an increasing need for text mining techniques in order to address the competition.

Large amounts of textual data could be collected as a part of a research, such as scientific literature, transcripts in the marketing and economic sectors, speeches in the field of political discourse, such as presidential campaigns and inauguration speeches, and meeting transcripts. Furthermore, online sources, such as emails, web pages, blogs/micro-blogs, social media posts, and comments, provide a rich source of textual data for research [1]. Large amounts of data are also being collected in semi-structured form, such as log files containing information from servers and networks. As such, text mining analysis is useful for both unstructured and semi-structured textual data [1].

Data mining and text mining differ on the type of data they handle. While data mining handles structured data coming from systems, such as databases, spreadsheets, ERP, CRM, and accounting applications, text mining deals with unstructured data found in documents, emails, social media, and the web. Thus, the difference between regular data mining and text mining is that in text mining the patterns are extracted from natural language text rather than from structured databases of facts [2]. Since all the written or spoken information can be represented in textual form, data mining requires all kinds of text mining tools when it comes to the interpretation and analysis of sentences, words, phrases, speeches, claims, adverts, and statements. This paper conducts an extensive analysis of text mining applications in big data analytics as used in various commercial fields and academic studies. While the vast majority of the literature deals with the optimization of a specific text mining technique, this paper seeks to summarize the features of all text mining methods, thereby summarizing the state-of-the-art practices and approaches in all the possible fields of application. It is centered around seven key applications of text mining in transcripts and speeches, meeting transcripts, and academic journal articles, as well as websites, emails, blogs, and social media networking sites; for each of these, we, respectively, provide a description of the field, their functionality, the most commonly used methods, the associated problems, and the related and relevant references.

The remaining sections of this paper are organized in the following manner. In the Section 2, we introduce the topic of text mining in transcripts and speeches. We explain the different classification techniques used in, for instance, the analysis of political speeches that classify opinions or sentiments in a manner that allows one to infer from a text or speech the ideology that a speaker most probably espouses. Furthermore, we explain the methods used in classifying transcripts and speeches and identify the shortcomings of these methods, which are primarily related to the behavioral nature of human beings, such as ironic or ideological behavior. In the Section 3, we take a closer look at blog mining, the dominance of news-related content in blogs and micro-blogging, and present the methods used in this area. Most of the methods applied in blog mining are based on dimensionality reduction, which is also found in other fields of text mining applications. Additionally, the relationship between blog mining and cybersecurity—which is an interesting and novel application of blog mining—is also covered in this section. In the Section 4, we analyze email mining and the techniques commonly used in relation to it. A very specific feature of email mining is its noisy data, which has been discussed in this section. Moreover, we explain the challenges to the identification of the content of the email body and how email mining is used in business intelligence. The web mining techniques that are used in screening and analyzing websites are studied in the Section 5. The features of a website, such as links, links between websites, anchor text, and html tags, are also discussed. Moreover, the difficulty of capturing unexpected and dynamically generated patterns of data is also explored. Additionally, the importance of pattern recognition and text matching in e-commerce is highlighted. In the Section 6, we present studies conducted on the use of Twitter and Facebook and explain the role of text mining in marketing strategies based upon social media, as well as the use of social media platforms for the prediction of financial markets. In Sections 7 and 8, we round up our extensive analysis of text mining applications by exploring the text mining techniques used for academic journal articles and meeting transcripts. Section 9 discusses the important issue of extract hidden knowledge from a set of texts and building hypotheses. Finally, in the concluding section, we highlight the advantages and challenges related to text mining and discuss its potential benefits to society and individuals.

2. Text Mining in Transcripts and Speeches

Text mining refers to the extraction of information and patterns that are implicit, previously unknown, and potentially valuable in an automatic or semi-automatic manner from immense unstructured textual data, such as natural-language texts [3].

There are two types of text mining algorithms: supervised learning and unsupervised learning (the two terms originated in machine learning methods). Supervised learning algorithms are employed when there is a set of predictors to predict a target variable. The algorithm uses the target's observed

values to train a prediction model. Support vector machines (SVMs) are a set of supervised learning methods used for classification and prediction. On the other hand, unsupervised learning methods do not use a target value to train their models. In other words, the unsupervised learning algorithms use a set of predictors (features) to reveal hidden structures in the data. Non-negative matrix factorization is an unsupervised learning method [4].

Transcripts are a written or printed version of material originally presented in another medium, such as in speeches. Therefore, the analysis of transcripts can be treated in the same manner as the analysis of speeches, as spoken words need to be pre-processed through, for instance, a voice-recognition API or manual transcription. Despite its extensive application in transcripts from other fields, such as marketing and political science, text mining as a technique in economics has historically been less explored. Bholat et al. [5] presented a comprehensive overview of the various text mining techniques used for research topics of interest to central banks for analyzing a corpus of documents, including, amongst others, the verbatim transcripts of meetings. Recently, three years of speeches, interviews, and statements of the Secretary General of Organization of the Petroleum Exporting Countries (OPEC) were analyzed using text-minin techniques [6].

The ideology, as a key factor affecting an individual's system of beliefs and opinions that controls their acts, is an important feature in text mining when it comes to political (or religious) textual data. Ideology provides the "knowledge of what-goes-with-what" [7] and shapes each individual's perception of any given issue [8]. However, the main issue in taking ideology as a feature for text and opinion mining is that, in many cases, the ideology of the speaker is not very clear, especially when it comes to politicians. To overcome the issue, one may use the texts with known ideological background and build a classification model to classify ideology behind a text, based on the textual data. Applying the trained classification model to a political text would help understand the ideology behind the speech or a text and, consequently, the opinion of a person.

Two approaches are extensively used in text mining: opinion classification and sentiment classification [9].

2.1. Opinion Classification

The main concern in opinion mining is to determine to what extent a text in-hand supports or opposes a specific subject. Although opinion mining is vastly used to analyze political texts, from speeches to short text on Twitter [8], it is very useful in other fields, too. For instance, one may use opinion mining to determine the opinion of the customers on features of a product, an audience's opinion on a movie, or to find the people's favorite asset in a market [10–12]. Most applications, in the context of political speeches, target the curation of general-purpose political opinion classifiers, given their potential and significant uses in e-rulemaking and mass media analysis [13–16]. The steps involved in the implementation of opinion mining are as follows [17]:

1. Determining text polarity to decide whether a given text is factual in nature (i.e., it unbiasedly describes a particular situation or event and refrains from providing a positive or a negative opinion on it) or not (i.e., it comments on its subject matter and expresses specific opinions on it), which amounts to the categorization of binary texts into *subjective* and *objective* [18,19].
2. Determining text polarity to decide if a given subjective text posits a positive or negative opinion on the subject matter [18,20].
3. Determining the extent of text polarity to categorize the positive opinion extended by a text on its subject matter as weakly positive, mildly positive, or strongly positive [21,22].

The literature related to opinion mining is growing see [23,24]. A highly beneficial source for opinion classification is Wordnet [25] by Princeton University, a lexical database of the English language containing nouns, verbs, adjectives, and adverbs grouped into 117,000 sets of cognitive synonyms (synsets), with each set expressing a distinct concept. A detailed description of Wordnet can be found in an article by Miller et al. [26].

2.2. Sentiment Classification

Sentiment classification is closely related to opinion mining and is mainly based on a technique called sentiment scoring. The basic idea behind the technique is to extract effective content from a text based on the appraisal, polarity, tone, and valence [27]. In order to build a sentiment score, one may use a set of predefined lists of terms with allocated quantitative weights for positive and negative connotations. Then, counting the positive and negative terms will get a score showing how much a text opposes or approves a given subject [28]:

$$\text{Sentiment Score} = \frac{\#\text{positive terms} - \#\text{negative terms}}{\#\text{all terms}}. \tag{1}$$

Taking the weights into account (if weights already exist):

$$\text{Sentiment Score} = \frac{\sum_i w_i^+ - \sum_j w_j^-}{\sum_i w_i^+ + \sum_j w_j^-}, \tag{2}$$

where w_i^+ is the sentiment weight for ith positive term, and w_j^- is the sentiment weight for jth negative term.

This measure is subsequently interpreted as a relative gap between positively and negatively connoted language. In a seemingly convenient manner, it ranges between -1 and $+1$, where a score of 0.5, for example, is interpreted as 50% points overweight for positively connoted language, implying a fairly positive sentiment guiding the text [27].

However, despite its strengths, such as implementation transparency, relevance, replicability, intuitiveness, and a high level of human supervisory, sentiment classification also bears some drawbacks, such as context dependence, which might hold a positive connotation in their original context (e.g., commercial reviews) but convey a negative tone in political contexts, or vice versa. Furthermore, estimating the positive and negative weights is not always straightforward.

According to Rauh [27], the more technically advanced literature has recently explored context-specific machine-learning approaches (e.g., Ceron et al. [29], Hopkins et al. [30], Oliveira et al. [31], and van Atteveldt et al. [32]). They also addressed its challenges, such as oversimplification, irony, and negation.

2.3. Functionality

One interesting application of opinion and sentiment classification is to use them for predicting someone's opinion or their system of beliefs and ideology based on their speech or written messages (e.g., text on social media, books, articles, etc.). Klebanov et al. [33] were the first researchers in the area of text classification to examine whether two people hold differing opinions or the same opinion but phrase it differently [34]. Moreover, they offered some insight into the conceptual structure that governs political ideologies, such as how these ideologies succeed in creating coherent belief systems, and determine (for the benefit of those who follow them) what goes with what. However, the results obtained in this manner provide a negligible amount of information regarding the structure of ideologies or the extent to which they are cohesive or convincing. In contrast, the studies conducted by Lakoff [35], Lakoff and Johnson [36], and Klebanov et al. [33] identify the underlying belief systems based on the cognitive structure and metaphors of liberal and conservative ideologies by employing an automatic lexical cohesion detector on Margaret Thatcher's 1977 speech for the Conservative Party Conference. The identification of the underlying belief systems requires the pre-processing of the text. Miner et al. [37] proposed a pre-processing method by removing stopwords (e.g., "the", "a", "an", etc.), prefixes (e.g., "re", "pre", etc.), and suffixes (e.g., "ing", "ation", "fy", "itis", etc.). Unifying words' spellings and typesettings (lower and upper cases) and correcting misspelled words is another step in their pre-processing scheme. The pre-processing will make the words normalized in the text and reduce the noise in unstructured text data. Sarkar et al. [38] described a classification algorithm based on a SVM, which allows for 80–89% accuracy.

Sentiment and opinion classification is used for classification of discussion treads and reviews, too. Lu et al. [39] used the opinion classification methods to automatically discover the opposed opinion and build an opposing opinion network for a social thread. In order to build the network, they analyzed the agree/disagree relations between posts in a social network platform (i.e., a forum). The sentiment and opinion classification methods are developed based on machine learning methods (e.g., SVM, neural networks, naive Bayes, maximum entropy, and stochastic gradient descent) to classify the large number of online reviews [40,41]. Kennedy and Inkpen [40] and Tripathy et al. [41] applied classification methods to classify the Internet Movie Database (IMDB) movie reviews, though they did not build a network of opinions on the movies.

A challenge in the classification of political speeches is that political speeches feature far fewer sentiment words—typically, adjectives or adverbs—that have been identified to be most indicative of opinions, as in the case of movie reviews. Instead, political speeches tend to express opinions in the choice of the nouns. Moreover, nouns that hold no political connotations in common usage may come across as heavily-laden with political intentions when expressed in the context of a specific debate [8].

Acharya et al. [34] used classification algorithms that, by comparing the performance of logistic regression, SVM, and naive Bayes models (NB), analyzes the speech of a given United States (U.S.) presidential candidate from 1996–2016 to predict the candidate's political party affiliation, as well as the region and year in which the speech was delivered. They found a superior performance of the logistic regression, followed by SVM and NB methods. These results are supported by the work conducted by Joachims [42], who found that logistic regression classifies a presidential candidate's speech as democratic or republican most accurately. Thus, predicting the candidate's political party and the year in which the speech was delivered is relatively easy, while predicting the location in which the speech was delivered, proves to be significantly more difficult.

2.4. Arguments Extraction

As another application, text mining is used to extract facts and arguments, specifically from political speeches and documents. An argument, in a certain context, usually consists of two main parts: a claim and a series of minor and major premises to support the claim. The premises are known and already proven facts. Argument extraction is closely related to opinion mining and belief classifications. Extracting the arguments from a large amount of textual data, not only helps building a knowledge base for a given subject or a task, it also helps to reexamine different arguments, with different common bases, and produce new ones in large scale. Extracting arguments requires distinguishing between the claims and the fact, which will result in extracting the arguments, along with the underlying facts. The big potential of argument extraction in political textual data has attracted many researchers in text mining. For instance, Sardianos et al. [43] proposed a supervised technique, based on conditional random fields, to extract arguments and their underlying facts. They applied the method to web pages containing news and tagged speeches. Florou et al. [44] applied a variate of argument extraction methods to Greek language social media to estimate the public support for an unannounced (unpublished or unfinished) policies. They took into account the structure of the sentences and discourse markers, like connectives, as well as the tense and mood of the verbal construction, and showed the importance of verbal construction in argument extraction. Goudas et al. [45] developed a two-stage approach to extract the arguments made by bloggers and others in social media related to the arguments made by policymakers, as well as new arguments in social media. The proposed model is applied to a vast number of Greek language social media contents. Their method has a high accuracy rate in extracting arguments and building relational links between arguments and policies. Lippi and Torroni [46] used the machine learning methods to extract arguments from seven party leaders' debates during the 2015 UK general election. Their results show the importance of using voice features, along with textual data, when it comes to extracting claims and facts from a speech.

2.5. Methods

Generally, the first step in text mining (after cleansing the textual data and reducing the noise) is to represent the text using a proper model. A common text representation model represents a document as a vector of features. The feature vector represents text with its frequent words or phrases, or grammatical structure of the sentences [47]. Most simple vector representations of text consider a text as a Bag-of-Words (BOW) or combination of BOW with Bag-of-Characters. More advanced versions look at the text as a Bag-of-Features [48]. Another common text representation method is to use a graph or a diagram to represent the relation of the words and segments in a document [49]. The diagram representation can be used to demonstrate the relation between terms in a text and use these terms to classify/categorize the text [48]. The next problem in text mining is to find a similarity measure and a classification function to properly classify the texts. One approach is to employ semantic similarity measures [50]. In addition to similarity measures, one may use machine or statistical learning models to classify textual data. The logistic regression is a binary classification algorithm that applies the logistic function as the hypothesis. The model subsequently locates the optimal θ that minimizes the associated cost function $J\theta$ that will then determine a separating sigmoid curve between the two classes [34]. SVM is a statistical classification method that was suggested by Cortes and Vapnik [51]. SVM, exploiting the structural risk minimization principle of computational learning theory, seeks a decision surface to categorize the training data points into two classes and forms decisions based on the SVMs which are identified as the only competent elements in the training set. According to Vinodhini and Chandrasekaran [52], SVMs separate the classes by building a margin in an effort to minimize the distance between each class and that margin.

NB models learn probabilities based on prior distribution across classes from the training data, under the assumption that all the features are independent; this specifically holds true when predicting a class based on training [34].

Yu et al. [8] tested the party classifiers for congressional speech data. They found that the classifiers which were trained on the house speeches are more efficient with processing senate speeches than vice versa and that the best overall classifier is SVM, which has equally weighted features.

In addition to the application of logistic regression, SVM, and NB, other statistical methods are also available for classification in natural language processing, such as maximum entropy and maximum likelihood [53], which use Candide, an automatic machine translation system developed by IBM, to test the performance of both the methods. These methods find a significant efficacy of maximum entropy techniques for performing context-sensitive modeling.

Other evidence for detectable patterns associated with ideological orientation in the political speech were found by Diermeier et al. [54], Evans et al. [55], and Laver et al. [56], as these studies achieve a high classification accuracy. Piryani et al. [57] presented an extensive scientometric analysis of the research work undertaken on opinion mining and sentiment analysis during 2000–2016.

Wilson et al. [22] created a system called the OpinionFinder, which performs a subjectivity analysis. Thus, it can automatically identify texts with opinions, sentiments, speculations, and other private states. OpinionFinder seeks to identify subjective sentences and to highlight the various aspects of subjectivity in these sentences, including the source (holder) of the subjectivity and words that are included in phrases expressing positive or negative sentiments. It encompasses four components:

1. An NB classifier that applies several lexical and contextual features to distinguish between subjective and objective sentences [58,59];
2. A component for identifying speech events (e.g., "stated" and "according to") and directing subjective expressions (e.g.,, "appalled" and "is sad");
3. A source identifier combining a conditional random field sequence tagging model [60] and extraction pattern learning [61] to determine the sources of the speech events and subjective expressions [62];

4. A component that applies two classifiers to identify the words contained in phrases that express positive or negative sentiments [63].

2.6. Shortcomings

Sentiment analysis, however, faces a predominant challenge with its classification of text under one particular sentiment polarity, whether positive, negative or neutral [24,64–67]. In order to solve this problem, Fang [68] proposed a general process for the categorization of sentiment polarity.

Another field of application is the detection of offensive language in the so-called hate speech, which refers to submitting to stereotypes to express and propagate an ideology of hate [69,70].

A key problem in speech recognition is that transcripts with high word error rate are obtained for documented speeches in poor audio conditions and spontaneous speech recorded in actual conditions, as pointed out by the NIST Rich Transcription Meeting program [71]. Recordings from Call centers and telephone surveys are of poor audio quality due to the use of cell phones and/or surrounding noise, unconstrained speech, variable utterance length, and various disfluencies, such as pauses, repetitions, and rectifications. Consequently, speech mining is extremely difficult on this type of corpora [72]. Camelin et al. [72] proposed a sampling and information extraction strategy as the solution to these problems. In order to evaluate the accuracy, as well as the representativeness, of the extracted information, they suggested several solutions based on the Kullback-Leibler divergence.

3. Blog Mining

Blogs allow authors to maintain entries that are continuing and arranged in reverse chronology for an audience that can interact with the authors through the comments section. Blogs can belong to a broad variety of genres, ranging from diaries of personal and mundane musings to corporate business blogs; however, they tend to be associated with more personal and spontaneous forms of writing. Social researchers have capitalized on blogs as a source of data in several cases, from performing content analysis related to gender and language use to determining ethnographic participation in blogging communities [73]. After the creation of the very first blog, Links.net, in 1994 [74], the internet became home to hundreds of millions of blogs. Due to the large numbers of existing blog posts, the blogosphere content may seem haphazardly and chaotic [75]. Consequently, effective mining techniques are required to aid in the analysis and comprehension of blog data. Webb and Wang [76] reviewed the general methodological options that are frequently used when studying blogs and micro-blogs; the options investigated included both quantitative and qualitative analyzes, and the study was undertaken in an effort to offer practical guidance on how a researcher can reasonably sift through them.

Many of the blog mining techniques are similar to those used for text and web documents; however, the nature of blog content may lead to various linguistic and computational challenges [77,78]. Current research in blog mining reflects the prominence of news or news-related content and micro-blogging. Blog mining, furthermore, overlaps with features of social media mining [79,80].

Apart from the text content, blogs also provide other information, such as details regarding the title and author of the blog, its date and time of publication, and tags or category attributes, among others. Similar to other social media data, blog content also undergoes changes over time. New posts are uploaded, novel topics are deliberated over, perceptions change, and new communities spring up and mature. Identifying and understanding the topics that are trending in the blogosphere can provide credible information regarding product sales, political views, and potentially attention-garnering social areas [81,82].

Methods in blog mining that have gained popularity over the years include classification and clustering [83], probabilistic latent sentiment analysis (PLSA) or latent Dirichlet allocation (LDA), mixture models, time series, and stream methods [80]. Existing text mining methods and general dimensionality reduction methods have been used by a number of studies [75,77,84–87] on blog mining; however, the analysis that can be undertaken with these methods is limited to mono- or bi-dimensional blog data, while the general dimensionality methods may not be effective in preserving information

retrieved from blogs [88]. Tsai [78] applied the tag-topic model for blog data mining. Dimensionality reduction was performed with the spectral dimensionality reduction technique Isomap to show the similarity plot of the blog content and tags. Tsai [88] presented an analysis of the multiple dimensions of blog data by proposing the unsupervised probabilistic blogger-link-topic (BLT) model to address the challenges in determining the parties most likely to blog about a specific topic and in identifying the associated links for a given blog post on a given topic and detect splog. The results indicated that BLT obtained the highest average precision for blog classification with respect to other techniques that used the blogger-date-topic (BDT), author-topic (AT), and LDA models. In the study of Tsai [89], the AT model based on the LDA was extended to the analyzes and visualization of blog authors, associated links, and time of publication, and a framework based on dimensionality reduction was suggested to visualize the dimensions of content, tags, authors, links, and time of publication. This study was the first to analyze the multiple dimensions of blogs by using dimensionality reduction techniques, namely multidimensional scaling (MDS), Isomap, locally linear embedding (LLE), and LDA, on a set of business blogs.

Sandeep and Patil [90], after conducting a brief review of the literature on blog mining, proposed a multidimensional approach to blog mining by defining a method that combines the blog content and blog tags to discern blog patterns. However, the proposed method can only be applied to text-based blogs.

Blogs can be categorized, influential blogs can be promoted, and new topics can be identified. It is also possible to ascertain perspectives or sentiments from the blogosphere through data mining techniques [81]. Although several solutions are available that can effectively handle information in small volumes, they are static in nature and usually do not scale up accurately owing to their high complexity. Moreover, such solutions have been designed to run once or in a fixed dataset, which is not sufficient for processing huge volumes of streamed data. In response to this issue, Tsirakis et al. [91] suggested a platform focusing on real-time opinion mining from news sites and blogs. Hussein [92] presented a survey on the challenges relevant to the approaches and techniques of sentiment analysis. Furthermore, the research of Chen and Chen [93] applied big data and opinion mining approaches to the analysis of investors' sentiments in Taiwan. First, the authors reviewed previous studies related to sentiment mining and selection of features; subsequently, they analyzed financial blogs and news articles available online for creating a public mood dynamic prediction model exclusively for Taiwanese stock markets by taking into account the views of behavioral finance and the features of financial communities on the internet.

The filtering of spam blogs is another predominant theme in blog mining, and it can considerably misrepresent any estimation of the number of blog posts made [78] and the evaluation of cybersecurity threats. Most intelligence analysis studies have focused on analyzing the news or forums for security incidents, but few have concentrated on blogs. Tsai and Chan [85] analyzed blog posts for identifying posts made under various categories of cyber security threats related to the detection of cyber attacks, cybercrime, and terrorism. PLSA was used for detecting keywords from various cyber security blog entries pertaining to specific topics. Along similar lines, Tsai and Chan [94] proposed blog data mining techniques for assessing security threats. They used LDA-based probabilistic methods to detect keywords from security blogs with respect to specific topics. The research concluded that the probabilistic approach can enhance information retrieval related to blog search and keyword detection. Recognition of cyber threats from open threat intelligence can prove beneficial for incident response in very early stages. Lee et al. [95] proposed a free web service for examining emerging cybersecurity topics based on the mining of open threat intelligence, which is dedicated to locating various emerging topics in cyber threats (i.e., nearly zero-day attacks) and providing possible solutions for organizations. The demonstration showed that with information collected from experts on Twitter and specific targeted RSS blogs, Sec-Buzzer promptly recognizes the emerging information security threats and, subsequently, publishes related news, technical reports, and solutions in time.

Applications of blog mining vary and, among others, include opinion mining for agriculture [96], prospective industrial technologies [97], decision support in fashion buying processes [98], detection of major events [99], retrieval of information regarding popular tourist locations, and travel routes [100], the summarization of popular information from massive tourism blog data [101], summarization of news blogs and detecting the copy and reproduced multi-lingual contents [102–104] and detecting the fake news [105].

4. Email Mining

Email is a convenient and common means of textual communication. It is also intrinsically connected to the overall internet experience since an email account is required for signing up for any form of online activity, including to create accounts for social networking platforms and instant messaging. The Radicati Group, Inc., executive summary of email statistics report 2012–2019 [106] predicts that the total number of business and consumer emails sent and received daily will exceed 293 billion in 2019; this statistic is forecasted to increase to over 347 billion by the end of 2023. To optimize the use of emails and explore its business potential, email mining has been extensively undertaken and has observed commendable progress in terms of both research and practice.

Email mining is similar to text mining since they both pertain to textual data. However, specific characteristics of email data separate it from text mining. To begin with, email data can be highly noisy. More specifically, it may include headers, signatures, quotations, or program codes. It may also carry extra line breaks or spaces, special character tokens, or spelling and grammar mistakes. Moreover, spaces and periods may be mistakenly absent from it. Hence, the email data needs to be cleaned in depth before high quality mining [107]. In addition, an email is a data stream targeted towards a specific user and the concepts or distributions of the target audiences of the messages may vary over time with respect to the messages received by that user. It is also problematic to obtain public email data for experiments due to privacy issues [108].

Emails contain links to a vast social network with data on the person or organization in charge, thus making email mining more resourceful [109]. During email mining, the links can be exploited for their content and a better understanding of behavior. On the other hand, the techniques currently in use are not able to effectively handle the vast amount of data. There is heterogeneity, noise, and variety, and mining techniques cannot be easily modified to adjust to big data environments [110]. Another challenge in email mining is data visualization, which makes decision-making considerably harder. Hidden information cannot be extracted or visualized due to the lack of scalable visualization tools. If the data is not presented more comprehensibly, visualization and decision-making also become difficult for the data miners [111].

Initial studies on email mining predominantly focused on the existing tools for personal collection management since large and diverse collections were not accessible for research use [112]. That changed, most notably with the Enron Corporation email collection [113]. However, the LingSpam corpus, compiled by Androutsopoulos et al. [114], was one of the first publicly available datasets.

A number of email mining studies have focused on people-related tasks, including name recognition and reference resolution, contact information extraction, identity modeling and resolution, role discovery, and expert identification, as well as the generation of access to large-scale email archives from multiple viewpoints by using a faceted search [115]. Moreover, some significant applications of email mining include tasks, such as filtering emails based on priority and identifying spam and phishing emails, as well as automatic answering, thread summarization, contact analysis, email visualization, network property analysis, and categorization [116,117].

Tang et al. [116] conducted a brief but exhaustive survey on email mining. The authors introduced the feature-based and social structure-based representation approaches, which are often performed in the pre-processing phase. Following this, they identified five email mining tasks—spam detection, email categorization, contact analysis, email network property analysis, and email visualization. Later, the commonly used techniques for each task were discussed. These included NB, SVMs, rule-based

and content-based models, and random forest (RF), as well as K-nearest neighbour (K-NN) classifiers (pertaining to the classification problem in the email content detection) and the K-means algorithm (pertaining to the semi-supervised clustering problem). The methods based on principal component analysis (PCA), LDA, and term frequency-inverse document frequency (TF-IDF) were presented as well.

The study by Mujtaba et al. [117] comprehensively reviewed 98 articles published between 2006–2016 on email classification from the Web of Science core collection databases and the Scopus database. In this study, the methodological decision analysis was performed in the following five aspects: (1) email classification application areas, (2) the datasets used in each application area, (3) feature space utilized in each application area, (4) email classification techniques, and (5) the use of performance measures.

Sentiment analysis of online text documents has been a flourishing field of text mining among researchers and scholars. In contrast to the content of public data, the real sentiment is often expressed in personal communications. Emails are frequently used for sending emotional messages that reflect deeply meaningful events in the lives of people [118]. On the other hand, sentiment analysis on large business emails could reveal valuable patterns useful for business intelligence [119,120]. The study of Hangal et al. [118] proposed the use of sentiment analysis techniques on the personal email archives of users to aid the task of personal reflection and analysis. The authors built and publicly released the Muse email mining system. The system helps users to analyze, mine, and visualize their own long-term email archives. Moreover, Liu and Lee [120] proposed a framework for email sentiment analysis that uses a hybrid scheme of algorithms, combined with K-means clustering and SVM classifier, and is to be applied to the Enron email corpus. The evaluation for the framework is conducted by comparing three labeling methods, namely, SentiWordNet, K-means, and polarity, and five classifiers, namely, SVM, NB, logistic regression (LR), decision tree (DT), and OneR. The empirical results indicated that the combined K-means and SVM algorithm achieved high accuracy compared to other approaches. In continuation of their previous studies, Liu and Lee [121] conducted sentiment clustering on Enron email data with a novel sequential viewpoint. This involved the transformation of sentiment features into a trajectory representation for implementing the trajectory clustering (TRACLUS) algorithm, along with the combination of sentiment temporal clustering, so as to discover sentiment flow in email messages in the topical and temporal distribution.

While insider threats in cybersecurity are often associated with malicious activities, insider threat is one of the most significant threats faced in business espionage [122]. Chi et al. [123], focused on the detection of insider threats by combining linguistic analysis and K-means algorithm to analyze communications, such as emails, to ascertain whether an employee meets certain personality criteria and to deduce the risk level for each employee. Soh et al. [124] focused on an aspect-based sentiment analysis that can provide more detailed information. Moreover, they presented a novel employee profiling framework equipped with deep learning models for insider threat detection, which is based on aspect-based sentiment and social network information. The authors evaluated the new presented framework, ASEP, on the dataset of the augmented Enron emails, and demonstrated that the employee profiles retrieved from ASEP can effectively encode the implicit social network information and, more significantly, their aspect-based sentiments.

The continued growth in the number of email users has led to a massive increase in spam emails. The global average of the daily spam volume for June 2019 was 459.40 billion, while the corresponding average of the daily (legitimate) email volume was 79.82 billion [125]. The large volume of spam emails moving through computer networks has a debilitating effect on the memory space available to email servers, communication bandwidth, CPU power, and user time. On the other hand, if we consider the fact that the majority of cyber attacks start with a phishing email [126] into accounts, there can be no doubt that phishing is a high-risk attack vector for organizations and even government agencies. Therefore, a predominant challenge in the email mining process is to identify and isolate spam emails.

Two general approaches are adopted for mail filtering: knowledge engineering (KE) and machine learning (ML) [127]. Spam filtering techniques based on knowledge engineering use a set of predefined rules. These rules are implemented to identify the basic characteristics of the email message. The ML techniques construct a classifier by training it with a set of emails called the training dataset. Several filtering methods based on ML have been extensively adopted when addressing the problem of email spam.

Bhowmick and Hazarika [128] presented an exhaustive review of some of the frequently used content-based email spam filtering methods. They mostly focused on ML algorithms for spam filtering. The authors studied the significant concepts, efforts initiated, effectiveness, and trends in spam filtering. They comprehensively discussed the fundamentals of email spam filtering, the changing nature of spam, and spammers' tricks to evade the spam filters of email service providers (ESPs). Moreover, they examined the popular machine learning techniques used in combating the menace of spam.

Dada et al. [129] examined the applications of ML techniques to the email spam filtering process of leading internet service providers (ISPs), such as Gmail, Yahoo, and Outlook, and focused on revisiting the machine learning techniques used for filtering email spam over the 2004–2018 period, such as K-NN, NB, Neural Networks (NN), Rough set, SVM, NBTree classifiers, firefly algorithm (FA), C4.5/J48 decision tree algorithms, logistic model tree induction (LMT), and convolutional neural network (CNN). Stochastic optimization techniques, such as evolutionary algorithms (EAs), have also been explored by Dada et al. [129], as the optimization engines are able to enhance feature selection strategies within the anti-spam methods, such as the genetic algorithm (GA), particle swarm optimization (PSO), and ant colony algorithm (ACO).

Most of relevant works on this topic classify emails using the term "occurrence" in the email. Some works, additionally, focus on the semantic properties of the email text. In the study conducted by Bahgat et al. [130], the email filtering was based on the introduction of semantic modeling to address the high dimensionality of features by examining the semantic attributes of words. Various classifiers were studied to gauge their performance in segregating emails as spam or ham experiments on the Enron dataset. Correlation-based feature selection (CFS) which was introduced as a technique for feature selection, improved the accuracy of RF and radial basis function (RBF) network classifiers, while CFS ensured the accuracy of other classifiers, such as SVM and J48.

A phishing attack that uses sophisticated techniques that direct online customers to a new web page that has not yet been included in the black-list is called a zero-day attack [131]. Chowdhury et al. [132] in their overview of the work on the filtering of phishing emails and pruning techniques, proposed a multilayer hybrid strategy (MHS) for the zero-day filtering of phishing emails that emerge during a separate time span, which uses the training data collected previously during another time span. MHS was based on a new pruning method, the multilayer hybrid pruning (MHP). The empirical study demonstrated that MHS is effective and that the performance of MHP is better than that of other pruning techniques.

In a newer approach aimed at studying the detection of phishing emails, Smadi et al. [133] discussed the relevant work on protection techniques, as well as their advantages and disadvantages. Moreover, they proposed a novel framework that combines a dynamic evolving neural network, based on reinforcement learning (RL), to detect phishing attacks in the online mode for the first time. The proposed model, phishing email detection system (PEDS), was also the first work in this field that used reinforcement learning to detect a zero-day phishing attack. NN was used as the core of the classification model, and a novel algorithm, called the dynamic evolving neural network, which used reinforcement learning (DENNuRL), was developed to allow the NN to evolve dynamically and build the best NN capable of solving the problem. It was demonstrated that the proposed technique can handle zero-day phishing attacks with high levels of performance and accuracy, while comparison with other similar techniques on the same dataset indicated that the proposed model outperformed the existing methods.

5. Web Mining

The World Wide Web (or the web) is at present a popular and interactive medium for disseminating information. The most commonly accessed type of information on websites is textual data, such as emails, blogs, social media, and web news articles. Web data differs from the data retrieved from other sources because of certain characteristics that make it more advantageous. In fact, website information is readily available to the public at large, is cost-effective in terms of access, and can be extensive with respect to coverage and the volume of data contained [134]. However, locating information on the web is a daunting and challenging task because of the immense volume of data and noise contained [135].

Etzioni [136] referred to web mining as the application of data mining techniques to automatically discover and extract knowledge in a website, while Cooley et al. [137] further highlighted the importance of considering the behavior and preferences of the users. Web mining has enabled the analysis of the increasing volume of data accessible on the web. Furthermore, it has indicated that conventional and traditional statistical approaches are inefficient in undertaking this task [138]. Besides, clean and consolidated data is closely connected to the quality and utility of the patterns discerned through these tools since they are directly dependent on the data to be used [139].

Web usage mining (WUM), web structure mining (WSM), and web content mining (WCM) are the three predominant categories of web mining [136,140,141]. WCM and WSM utilize the primary web data, while WUM mines the secondary data [136].

WCM adopts the concept and principles of data mining to discover information from the text and media documents [142]. WCM mainly focuses on web text mining and web multimedia mining. WSM emphasizes the hyperlink structure of the web to link the different objectives together [143]. A typical web graph is structured with web pages as nodes and hyperlinks as edges, establishing a connection between two related pages. WSM primarily works on link mining, internal structure mining and URL mining. In addition, WSM can be used for categorizing web pages and is useful for gathering information, such as that pertaining to the similarities and relationships between different websites. The typical applications of WSM are (a) link-based categorization of web pages, (b) ranking of web pages through a combination of content and structure, and (c) reverse engineering of website models [144]. Link-based classification pertains to the prediction of a web page category, which is based on the words on the page, links existing between the pages, anchor text, HTML tags, and other potential attributes on a web page. WUM is the process of applying data mining techniques to the discovery of usage patterns from the web data [145]. When a user interacts with a website, web log data is generated on a web server in the form of web server log files. Different types of usage log files, such as access log, error log, referrer log, and agent log, are created on a server [146]. Web logs are the type of data that prove the most resourceful when performing a behavioral analysis on a website user [137]. Web usage mining consists of three phases: (1) pre-processing, (2) discovery of usage patterns, and (3) analysis of the pattern. Typical applications are (a) the ones based on user modeling techniques, such as web personalization, (b) adaptive web sites, and (c) user modeling [144]. However, this personalization process that contains rebuilding a user's session has raised important legal and ethical concerns. Velásquez [139] adopted an integrative approach based on the distinctive attributes of web mining to identify the harmful techniques.

Analyzing the patterns generated from a typical web user's complex behavior is a daunting task since, most of the time, a user is responsible for the spontaneous and dynamic generation of patterns of data [147,148]. The exploration of the web for outliers, such as noise, deviation, incongruent observations, peculiarities, and exceptions, has received attention in the mining community. Chandola et al. [149] provided a general and broad overview of the extensive research conducted on anomaly detection techniques, spanning multiple research areas and application domains, including web applications and web attacks. Gupta and Kohli [148,150,151] made experimental attempts to identify outliers in regression algorithm outputs by using web-based datasets. In fact, various regression algorithms are extensively adopted by several online portals operating in varying application domains,

especially e-commerce websites [148]. Specifically, Gupta and Kohli [151] formulated a framework with the help of ordered weighted operators (OWA) as a multicriteria decision-making (MCDM) problem. The results proved that the proposed framework can aid in considerably reducing the outliers; however, its testing was restricted to a static purpose and a small dataset and the data were scattered for over a year. This work was an extension of an earlier study by Gupta and Kohli [150] in which a small experiment was conducted on a web dataset through the application of an ordered weighted geometric averaging operator. A recent study by Gupta and Kohli [148] detected outliers based on the principle of multicriteria decision-making (MCDM) and utilized ordered weighted operators for the purpose of aggregation.

On a daily basis, news websites feature an overwhelming number of news articles. While several text mining techniques can be applied to web news articles, the constantly changing data characteristics and the real-time online learning environment can prove to be challenging.

Two recent studies, conducted by Iglesias et al. [152] and Za'in et al. [153], proposed a different approach based on evolving fuzzy systems (EFS). It allows the updating of the structure and parameters of an evolving classifier, aids in coping with huge volumes of web news, and enables the processing of data online and in real time, which is essential in real-time web news articles. Iglesias et al. [152] developed a web news mining based on eClass0 classifier, while Za'in et al. [153] proposed a web news mining framework built on fuzzy evolving type-2 classifier (eT2Class), which outperforms other consolidated algorithms.

With the effective use of e-commerce, the internet increases the accessibility of customers from all over the world without having to deal with any marketplace restrictions. Web mining research is emerging in many aspects of e-services with the aim of improving online transactions and making them more transparent and effective [154]. The owners of e-commerce websites depend considerably on the analysis and summarization of customer behaviors so as to invest efforts towards influencing user actions and optimizing the success metric. The application of web mining techniques on the web and e-commerce for the sake of improving profits is not new, and a significant amount of research has been conducted in this field, especially pertaining to usage data. Recently, Dias and Ferreira [155] proposed an all-in-one process, improved by the crossing of data secured from diverse sources, for collecting and structuring data from an e-commerce website's content, structure, and users. Finally, they presented an information model for an e-commerce website which contained the recorded and structured information resulting from the intersection of various sources and tasks for pattern discovery. Moreover, Zhou et al. [156] proposed three new types of automatic data acquisition strategies, based on web crawlers and the Aho-Corasick algorithm, to improve the text matching efficiency by considering the Chinese official websites for agriculture, the wholesale market websites of agricultural products, and websites for agricultural product e-commerce.

In the current era of vibrant electronic and mobile commerce, the financial transactions conducted online on a daily basis are massive in number, which creates the potential for fraudulent activity. A common fraudulent activity is website phishing, which involves creating a replica of a trustworthy website for deceiving users and illegally obtaining their credentials. A report published by Symantec Corporation Inc. [157] substantiated that the number of malicious websites detected rose by 60% in 2018 with respect to 2017.

The phishing phenomena, which mostly focused on web-based phishing detection methods than email-based detection methods, were discussed in detail by the study of Mohammad et al. [158], which provided a comprehensive evaluation of the blacklist-based, whitelist-based, and the heuristics-based detection approaches. The study concluded that, despite only heuristics-based detection approaches having the ability to recognize these websites, their accuracy may reduce considerably in case of change in the environmental features. A successful phishing detection model should also be adept at adapting its knowledge and structure in a continuous, self-structuring, and interactive manner in response to the changing environment that is characteristic of phishing websites. Yi et al. [159] proposed a class of deep neural network, namely the deep belief model (DBN), to detect web phishing.

They evaluated the effectiveness of the detection model on DBN based on the true positive rate (TPR) with different parameters. The TPR was found to be approximately 90%.

Diverse disciplines have been interested in and have extensively undertaken the analysis of human behavior. Therefore, a broad theoretical framework is available with remarkable potential for application in other areas, particularly in the analysis of web user browsing behavior. With respect to web user browsing behavior, a prominent source of data is web logs that store every website visitor's actions [160]. A recent study by Apaolaza and Vigo [161] addressed the challenges of mining web logs and proposed a set of functionalities into workflows that addresses these challenges. The study indicated that assisted pattern mining is perceived to be more useful and can produce more actionable knowledge for discovering interactive behaviors on the web. The requirement for more accurate and objective data for describing the navigation and preferences of web users led the researchers to study a combination of different data sources, from web and biometric data to traditional WUM research or experiments. Slanzi et al. [162] provided an extensive overview of the biometric information fusion applied to the WUM field.

6. Social Media

With the advent of social media, information related to various issues started going viral. Dealing with this flow has become an indispensable societal daily routine [163]. Moreover, social media creates new ways for people from various communities to engage with each other [164]. Social media is a perfect platform for the public to transfer opinions, thoughts, and views on any topic in a manner that significantly affects their opinions and decisions. Many companies simultaneously analyze the information available on social media platforms to collect the opinions of their customers and implement market research. Additionally, social media has started attracting researchers from several fields, including sociology, marketing, finance, and computer [24].

6.1. Twitter

Twitter is a social media platform where users can share their opinions, follow others, and comment on their opinions. In recent years, several researchers have focused on Twitter. With over 140 million tweets being posted in a day, Twitter serves as a valuable pool of data for many researchers [165]. Studies on topics ranging from the prediction of box office results of a movie to the changes in the stock market are based on Twitter data. Nisar and Yeung [166] collected a sample of 60,000 tweets made over a six-day period before, during, and after the local elections in the United Kingdom to investigate the relationship between their content and the changes in the London FTSE100 index [166]. Similarly, many other researchers use the information available on Twitter to make stock market predictions [167–175]. Öztürk and Ayvaz [163] studied Turkish and English tweets for evaluating their sentiments towards the Syrian refugee crisis and found that Turkish tweets are remarkably different from English tweets [163]. A study on the Arabic Twitter feed is proposed by Alkhatib et al. [176] with the objective of offering a novel framework for events and incidents management in smart [176]. Gupta et al. [177] presented a research framework to examine the cybersecurity attitudes, behavior, and their relationship by applying sentiment analysis and text mining techniques on tweets for gauging people's cybersecurity actions based on what they say in their texts.

Regarding tourist sentiment analysis, Philander and Zhong [178] expounded on the application of tourist sentiment series from Twitter data for building low-cost and real-time measures of hospitality customer attitudes/perceptions.

Twitter data has also been studied by researchers from various fields to analyze (a) the Twitter usage behaviors of journalists [179] and cancer patients [180], (b) the sentiments of political tweets during the 2012 U.S. presidential election [181], (c) the effects of Twitter on brand management [182] and a given smartphone brand's supply chain management [183], (d) the opinions held by people on the issue of terrorism [184], (e) the social, economic, environmental, and cultural factors pertaining to

the sustainable care of both the environment and public health which most concern Twitter users [185], and (f) the tweet posting comments of academic libraries [186].

6.2. Facebook

Facebook is an American social media platform that is considered to be one of the biggest technology companies besides Amazon, Apple, and Google. According to Social Times, Facebook has 1.59 billion monthly active users. The Pew Research Center [187] determined that Facebook is the most extensively used social media platform. Facebook allows its users to express their thoughts, views, and ideas in the form of comments, wall posts, and blogs [187].

Kim and Hastak [188] analyzed Facebook data during the 2016 Louisiana flood, when parishes in Louisiana used their Facebook accounts to share information with people affected by the disaster. They discussed the critical role played by social media in emergency plans with the aim of helping emergency agencies in creating better mitigation plans for disasters [188].

In the context of business, companies need to monitor customer-generated content, not only on their personal social media page but also on the page of their competitors, to increase their competitive advantages. In this regard, He et al. [189] apply text mining to the Facebook and Twitter pages of three of the largest pizza chains in the U.S. pizza industry in order to help businesses in utilizing social media knowledge for decision-making [189].

Salloum et al. [190] classified the Facebook posts of Arabic newspapers through different text mining techniques. They found that the UAE is a country that shares the most number of posts on Facebook and also that videos are the most attracting part of the Facebook pages of Arabic newspapers [190].

Text mining on Facebook is also used to help institutions with their marketing strategies. Al-Daihani and Abrahams [191] implemented a text analysis on the Facebook posts of the academic libraries of the top 100 English-speaking universities. Their findings can be applied by academic libraries to develop their marketing, engagement, and visibility strategies.

6.3. Other Social Media Platforms

Text mining on social media is also utilized to improve transport and tourism planning. In this regard, Serna and Gasparovic [192] conducted a study on transportation modes using TripAdvisor comments and proposed a dashboard platform with graphical items that analyzes this data. This dashboard would facilitate the results collected from social media and its effect on tourism would be discussed [192]. Furthermore, Sezgen et al. [193] investigated the primary drivers of customer satisfaction and dissatisfaction of both full-service and low-cost carriers and of economy and premium class cabins using TripAdvisor passenger reviews for fifty (50) airlines. Text mining in social media platforms has been used to automatically rank different brands and make recommendations. Suresh et al. [194] applied opinion mining to the real life reviews from Yelp. They used the reviews given by restaurants' customers to build a recommendation list. Saha and Santra [195] applied a similar idea to textual feedback from Zomato.

Existing literature on text mining on social media have predominantly discussed English texts and semantics [196] since most of the available packages have been developed for English-speaking users. However, several studies have focused on Chinese social media and semantics. Liu et al. [197] used discussion forums related to the Chinese stock market, namely the East Money forum, and opinion classification to predict stock volatilities [197].

Chen et al. [198] focused on Sina Weibo, a Chinese social media platform, to predict stock market volatilities. Moreover, they used the deep recurrent neural network [198].

A study by Liu et al. [199] sought to assess social media effects in the big data era. They used Chinese platforms, such as Hexun and Sina Weibo, and considered the stock index from 77 corporate companies in Shanghai and Shenzhen. The experimental results highlighted the positive relationship between the trading volumes/financial turnover ratios and the media activities [199].

The axis of the work of Zhang et al. [200] is Xueqiu, a Chinese platform that is similar to Twitter but specifically for investors. In their study, they classified the tweets by polarity, implementing the naive Bayes network, and predicted the stock price movements by using SVM and the perceptron network [200].

Recently, Pejic-Bach et al. [201] applied text mining on publicly accessible job advertisements on LinkedIn—one of the most influential social media networks for business—and developed a profile of Industry 4.0 job advertisements.

7. Published Articles

The enormous number of scientific publications provides an extremely valuable resource for researchers; however, their exponential growth represents a major challenge. On the other hand, a literature review is an essential component of almost any research project. Text mining enhanced the review of academic literature, and more papers are being published over the years using this technique. Text mining techniques can identify, group, and classify the key themes of a particular academic domain and highlight recurrence and popularity of topics over a period of time.

In the field of business, management, and information technology contexts, Moro et al. [202] performed topic detection on 219 articles between 2002 and 2013 through text mining when detecting terms pertaining to business intelligence in banking literature. They used the Bayesian topic model LDA and a dictionary of terms to group articles in several relevant topics. A similar study by Amado et al. [203], based on the study of Moro et al. [202], outlined a research literature analysis based on the text mining approach over a total of 1560 articles framed in the 2010–2015 period with the objective of identifying the primary trends on big data in marketing. Furthermore, Moro et al. [204] summarized the literature collected on ethnic marketing in the period 2005–2015 using LDA.

Cortez et al. [205] focused on analyzing 488 research articles published on a specific journal within a 17-year timeline on the domain of expert systems. The authors adopted LDA and followed a methodology similar to that applied by Moro et al. [202] and Moro and Rita [206] for branding strategies on social media platforms in the hospitality and tourism field. Guerreiro et al. [207] used the topic model cluster algorithm Correlated Topic Model (CTM), which is based on LDA, to conduct an analysis of 246 articles published in 40 different journals between 1988 and 2013 on the subject of cause-related marketing (CRM). The study revealed the most discussed topics on CRM. The study of Loureiro et al. [208] explored a text-mining approach using LDA to conduct an exhaustive analysis of 150 articles on virtual reality in 115 marketing-related journals, indexed in Web of Science. Galati and Bigliardi [209] implemented text mining methodologies for conducting a comprehensive literature review of Industry 4.0 to identify the main overarching themes discussed in the past and track their evolution over time.

Literature review articles on text mining have also recently emerged in the field of operations management, thus providing a framework for identifying the predominant topics and terms in the field. Guan et al. [210] used latent semantic analysis (LSA) to identify the core areas of production research based on the abstracts of all articles published in a specific journal since its inception and revealed how the focus extended on topics has evolved over time. Demeter et al. [211] applied two text mining tools on 566 papers in 12 special issues of a specific journal between 1994 and 2016 to gather a comprehensive review of the entire field of inventory research.

Literature reviews within several domains have also benefited from text mining. Grubert [212] investigated the Life Cycle Assessment (LCA) literature by applying unsupervised topic modeling to more than 8200 environment-related LCA journal article titles and abstracts published between 1995 and 2014. Yang et al. [213] mined 1000 abstracts from the Google Scholar database for search results for technology infrastructure of solar forecasting, classified the concepts of solar forecasting on the full texts of 249 papers from Science Direct, and also undertook the keyword analysis and topic modeling on six handpicked papers on emerging technologies related to the subject. Moro et al. [214] performed text mining over the whole textual contents of papers, excluding only the references and

authors' affiliations, published in a tourism-related journal from 1996 to 2016. In the field of agriculture, Contiero et al. [215] analyzed through text mining the abstracts of 130 peer-reviewed papers that were published between 1970 and 2017 dealing with the pain issue in pig production and its correlation with the welfare in pigs.

Text mining was further employed as a tool in literature review-based studies in public health and medical sciences for various key themes, such as the adolescent substance and depression [216], cognitive rehabilitation and enhancement through neurostimulation [217], the protein factors related to the different cancer types [218], and diseases and syndromes in neurology [219].

Text mining methods are also employed to extract metadata from published articles. Kayal et al. [220] developed a method to automatically extract funding information from scientific articles. Yousif et al. [221] developed a model based on deep learning to extract the purpose of a citation to an article. Coupling this information with the number of citation and the applications of the articles results gives good measure to evaluate the efficiency of the funding.

However, text mining comes under the microscope of copyright, contracts, and licenses. Invoking the fundamental principles of copyright in the context of new technologies, Sag [222] explains that copying expressive works for non-expressive purposes should not be considered as infringement and should instead be labeled as fair use. In his article, Sag deals with the U.S.'s current legal framework and the feasibility of its adjustment to the text data mining processes, especially in the aftermath of the decisions in the cases of Authors Guild v. HathiTrust and Authors Guild Inc. v. Google. Recently, the European's parliament voted in favor of a copyright exception (Directive (EU) 2019/790) on text and data mining for research purposes, as well as for individuals and institutions with legal access to protected works [223].

8. Meeting Transcripts

Most of the time, the meeting transcripts are too long, making reading and analyzing the core content infeasible; therefore, providing a framework that can extract the keywords automatically from the meeting transcripts is instrumental. Towards this end, Sheeba and Vivekanandan [224] proposed a model in which the keywords and key phrases are extracted from meeting transcripts. They claimed that the difficulty of this work is tied to the occurrence of synonyms, homonyms, hyponymy, and polysemy in the transcripts. Keywords were extracted by using the MaxEnt and SVM classifiers, and the extraction of bigram and trigram keywords was ensured through the N-gram-based approach.

The way meeting transcripts are written varies in terms of their style and details compared to the written text style. Liu et al. [225] presented a list with the differences that could negatively impact a keyword extraction system, such as the low lexical density, the lack of a perfect structure, the poor structure, and the varied speaking styles and word usage of a multiplicity of participants.

Liu et al. [225] extended the previous work of [226] by proposing a supervised framework for the extraction of keywords from meeting transcripts based on various features, such as decision-making sentence features, speech-related features, and summary features, that reflect the meeting transcripts more efficiently. The authors conducted experiments using the ICSI meeting corpus for both human transcripts and different automatic speech recognition (ASR) outputs, and they showed that the method suggested outperforms the TF-IDF weighting and a predominant state-of-the-art phrase extraction system.

Song et al. [227] published two different articles on the extraction of keywords from meeting transcripts. The authors proposed a just-in-time keyword extraction method by considering two factors that make their work different from that of others. These factors are (1) the temporal history of preceding utterance, which gives more importance to recent utterance, and (2) the topic relevance, which considers only the preceding utterances that are relevant to the current ones. Their method was applied on two English and Korean datasets, including the National Assembly transcripts in Korean and the ICSI meeting corpus. The results indicated that including these factors can enhance

keyword extraction. Furthermore, in a more recent study [228], they added the participant factor to their graph-based keyword extraction method.

Xie end Liu [229] applied a framework on the ICSI meeting corpus to summarize meeting transcripts. In this framework, only the noteworthy sentences are selected to form the summary. Therefore, to specify whether a sentence should be included in the summary or not, they used supervised classification. Moreover, various sampling methods were implemented for avoiding the problem of imbalanced data.

Sharp and Chibelushi [230] presented an algorithm for the text classification of meeting transcripts. This study focused on the analysis of spoken meeting transcripts of discussions on software development, with the aim to determine the topics discussed in these meetings and, thereby, extract the decisions made and issues discussed. The authors suggested an algorithm that is appropriate for segmenting meeting transcripts that are spoken by combining semantically complex lexical relations with speech cue phrases to build lexical chains in determining topic boundaries. They argue that the results can help project managers in avoiding rework actions.

9. Knowledge Extraction

As the number of digital documents has grown exponentially, almost on any topic, automated knowledge extraction methods have become popular in many sections, from science to marketing and business. Many text mining techniques deal with extracting knowledge authors represented in their texts, e.g., extracting arguments [43–46] and extracting opinions [13–16]. Knowledge extraction is usually concerned with finding and extracting the key elements from textual data (e.g., articles, short notes, tweets, blogs, etc.) to extract hidden knowledge or build new hypotheses. For instance, a set of articles by an author, (or group of authors) contains knowledge about writing style, reasoning methodology, etc. [231–233]. For another example, the set comments on the problem may contain information about different aspects of an issue and possible (proposed) solutions [234]. Extracting such knowledge requires the modeling of the relational structure of textual data, i.e., the relation between words, paragraphs, and other fractions of texts. Networks, as a powerful tool, have been used to demonstrate relational structure of textual data in many knowledge extraction applications.

The idea of using networks to represent the structure of a given text (or set of texts) has been a practical one in literary analysis. For instance, Amancio et al. [231] used a complex network to quantify characteristics of a text (e.g., intermittency, burstiness, and words' co-occurrence) to authorship attribution. Although the method successfully classified eight authors from the nineteenth century, they concluded that the accuracy of the results may depend on the text database, features extracted from the network, and the attribution algorithm. They suggest that different algorithms and features should be tested before engaging the real application. Amancio [233] employed the same idea and combined it with Fuzzy classification and traditional stylometry methods to classify texts based on authors' writing style. The results show improvement in authorship attribution and genre identification when a hybrid algorithm is used to classify texts. Although Amancio et al. [231] shade light on authorship attribution using networks and made it possible to think about automated author attribution, their study did not take the structural changes into account. It is very common among authors to change style over time due to change in socio-economic characteristics, personal changes, etc. Amancio et al. [232] employed a similar idea of using networks to detect and model the literary movements through time. The used the books published between 1590 to 1922 to draw the literary movements in five centuries.

Nuzzo et al. [235] developed a set of tools to discover new relations between genes and between genes and diseases so it can be used to build more likely hypotheses on gene-disease associations. In order to find new possibilities, they applied an algorithm with the following steps on the abstracts of published articles from "PubMed" databases:

1. Extract concepts on terms describing genes and diseases from abstracts.
2. Derive genes-disease annotation.

3. Use similarity metrics to demonstrate the relevance between genes, which measures the terms shared between genes to identifies the possible relations.
4. Summarize the resulting annotation network as a graph.

The method shows its power to identify new possible gene-disease relations and builds possible hypotheses, as well as extracts the existing knowledge from the abstracts. Although their method is effective, it cannot be easily used for other applications since it uses an already existing structured knowledge base (e.g., Unified Medical Language system) to extract concepts. In many applications, such a knowledge base either does not exist or is very limited.

Wang et al. [234] used a multilayer network structure to systematically characterize a social network conversations based on their contents. The study targets a large set of Twitter messages exchanged between people with Eating Disorder (ED) to determine the type of content discussed on the online community, determine the flow pattern for different types of contents, and show how the contents are correlated. Their use of multilayer network provides the ability to represent the nature of discussion and multiplex interactions (presenting each topic in a separated layer), as well as to incorporate multidimensional information. Furthermore, analyzing the multilayer networks for a period of time reveals the structural changes in connections, correlation between topics, etc. Their results show that engagement in pro-recovery and pro-ED discussion is highly correlated, as well as the number of entries and exits to a communication when there is pro-ED sharing. The flexibility of the model developed by Wang et al. [234] provides a powerful tool for extracting knowledge hidden in social network conversations (messages, comments, etc.).

10. Conclusions

New technologies have facilitated access to immense quantities of digital text, recording an ever increasing share of human interaction, communication, and culture [236]. Text mining provides a framework to maximize the value of information within large quantities of text; thereby, the use of text mining technologies has increased steadily in recent years and has become highly diverse.

This study has summarized the academic research efforts on text mining and its applications by examining the published literature developed over the recent past few years. Figure 1 shows the methods and applications discussed in this study. More than 200 academic journal articles on the subject were included and discussed in this review, alongside the state-of-the-art text mining approaches used for analyzing transcripts and speeches, meeting transcripts, and academic journal articles, as well as websites, emails, blogs/micro-blogs, and social media networking sites across a broad range of application areas.

In practice, text mining enables the efficient exploitation of textual data on a broad range of real-world applications, such as (a) supporting large companies in faster and better decision-making by providing insights on the performance of marketing/sales strategies, enhancing customer experience, monitoring and enhancing the product/service, and gaining better customer engagement, (b) analyzing documents and verbatim transcripts in the economics sector, (c) analyzing political discourse streams that may provide valuable insights into critical discourse analysis, (d) creating more reliable and effective filtering methods for emails and websites, and (e) identifying relationships between users and certain products for social media purpose, as well as examining opinions on particular topics or sentiments on certain events. At the same time, by mining immense amounts of information in scientific literature, researchers can discover patterns and links between resources that cannot be detected through usual human viewing and reading, provide more meaningful answers to complex research questions, and even support scientific discovery in various domains.

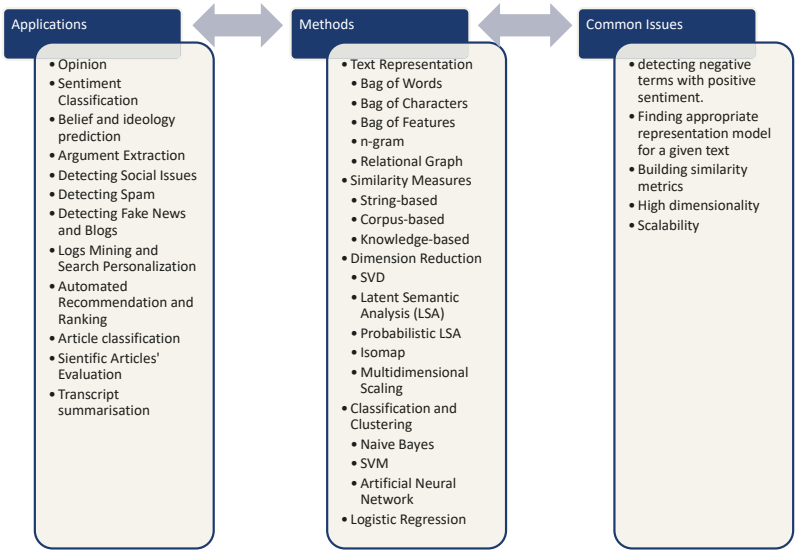


Figure 1. The methods and application discussed in this study.

There is a push, however, towards applications of text mining technologies on emerging crucial issues. One of the major serious issues is the relatively recent phenomenon of cybercrime [237–240] with strong impact on citizens, societies, and economies [241–244]. There are several ways in which text mining can be utilized for security analytics. Emails can be analyzed for discerning patterns in words and phrases, which may help identify a phishing attack. Websites can be scraped and analyzed to locate trends in themes that are related to security, such as the latest botnet threats, malware, and other Internet hazards [1]. More interestingly, social media offers a repository for intelligence-led policing operations; thereby, the law enforcement community is increasingly turning to social media monitoring to prevent and investigate crimes. Techniques, such as text mining, NLP, and sentiment analysis, provide a varied toolset that may assist in this direction [245]. Without pausing to address the approaches of previous studies regarding cybersecurity here in this paper, it is relevant to note that there is no universally agreed upon classification scheme that would contribute towards our understanding of cybercrime and serve as a useful tool for cybercrime stakeholders [246]. Recently, Donalds and Osei-Bryson [246] designed a new cybercrime classification scheme. Nevertheless, they also pointed out, the use of text mining and artificial intelligence technologies on this new ontology should be explored.

Another emerging, serious issue is the identification and detection of the widespread misinformation on social media and websites. In fact, the use of mega-platforms, such as Facebook and Twitter, as vectors for widespread misinformation spreading, e.g., during tragedies, national crises, or political campaigns, has been the subject of collective anxiety and a growing field of research [247–251]. Moreover, while one of the most beneficial values of text mining in big data analytics for businesses and governments is derived from the monitoring of human behavior and its predictive potential, the massive collection, instantaneous transmission, and combination and reuse of personal information for unforeseen purposes have placed new strains on strictly following the principles of data protection, which calls for a thorough consideration of their applications [252]. Serious ethical concerns and legal aspects have been raised when text mining is executed over data of a personal nature [139,253,254].

As it was noted earlier in this paper, there are pertinent challenges to the text mining process. First, the problem of ambiguity that the natural language faces is an issue. It can also be argued that what are conventionally referred to as languages exhibit immense internal variability across geographical and social space [255]. Moreover, many textual data sources are rife with abbreviations, acronyms, and specialized language. Second, the world of emails and online social networking sites can be very noisy. It may contain a large number of non-words, unknown words, and grammatically poor or incoherent sentences, as well as bots and trolls. Furthermore, text mining also carries limitations with respect to copyright, contracts, and licenses.

Another challenge in text mining arises when the method is employed in big textual data analysis. Since the size of big textual data rapidly grows, the text mining methods should be compatible with scalable data platforms. In other words, the employed text mining methods should have the ability to reduce the dimension of analyzed data and/or be compatible with the distributed computational systems and databases [256,257].

In summary, text mining carries immense potential as a tool for retrieving and analyzing large-scale and complex data and also allows spanning across a range of fields, disciplines, cultures, and languages. Not only are the cutting-edge of text mining technologies making significant improvements in terms of performance and accuracy within the framework of artificial intelligence and deep learning, mining in big data analytics is an evolving field, hence its having immense potential to advance science, encourage business growth in multiple industries, and ensure job growth. Moreover, text mining professionals are increasingly becoming high in demand. Furthermore, text mining may have the power to deliver significant insights to society and individuals, especially with respect to public health [258,259], healthcare [260,261], and education [262–265], and help evaluate social issues, such as crime (including cybercrime) [245,266,267], child abuse [268], and poverty [269]. Nevertheless, actions must be taken in time to efficiently solve the legal, ethical, and privacy concerns contained in the use of personal data.

Author Contributions: Conceptualization, H.H., C.B., S.U., M.T.M., M.R.Y.; investigation, H.H., C.B., S.U., M.T.M., M.R.Y.; writing–original draft preparation, H.H., C.B., S.U., M.T.M., M.R.Y.; writing–review and editing, H.H., C.B., S.U., M.T.M., M.R.Y.; supervision, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Talabis, M.R.M.; McPherson, R.; Miyamoto, I.; Martin, J.L.; Kaye, D. Security and text mining. In *Information Security Analytics*; Talabis, M.R.M., McPherson, R., Miyamoto, I., Martin, J.L., Kaye, D., Eds.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 123–150. [CrossRef]
2. Hearst, M.A. Text Data Mining. In *The Oxford Handbook of Computational Linguistics*; Mitkov, R., Ed.; Oxford University Press: Oxford, UK, 2005; pp. 616–662. [CrossRef]
3. Dumais, S. Using SVMs for text categorization, Microsoft research. *IEEE Intell. Syst. Mag.* **1998**, *13*, 18–28.
4. Guduru, N. Text Mining with Support Vector Machines and Non-Negative Matrix Factorization Algorithms. Ph.D. Thesis, University of Rhodes Island, Rhodes Island, Greece, 2006.
5. Bholat, D.; Hansen, S.; Santos, P.; Schonhardt-Bailey, C. *CCBS Handbook No. 33, Text Mining For Central Banks*; Bank of England: London, UK, 2015.
6. OPEC Bulletin. Language Lessons, July–August 2019. Available online: https://www.opec.org/opec_web/static_files_project/media/downloads/publications/OB07_082019.pdf (accessed on 1 January 2020)
7. Poole, K.T. Changing minds? Not in Congress! *Public Choice* **2007**, *131*, 435–451, doi:10.1007/s11127-006-9124-y. [CrossRef]
8. Yu, B.; Kaufmann, S.; Diermeier, D. Classifying party affiliation from political speech. *J. Inf. Technol. Polit.* **2008**, *5*, 33–48, doi:10.1080/19331680802149608. [CrossRef]

9. Esuli, A. A Bibliography on Sentiment Classification. 2006. Available online: http://liinwww.ira.uka.de/bibliography/Misc/Sentiment.html (accessed on 27 June 2019).
10. Dave, K.; Lawrence, S.; Pennock, D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (WWW2003), Budapest, Hungary, 20–24 May 2003; pp. 519–528, doi:10.1145/775152.775226. [CrossRef]
11. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2004), Seattle, WA, USA, 22 August 2004; pp. 168–177, doi:10.1145/1014052.1014073. [CrossRef]
12. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), Philadelphia, PA, USA, 6–7 July 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; pp. 79–86, doi:10.3115/1118693.1118704. [CrossRef]
13. Agrawal, R.; Rajagopalan, S.; Srikant, R.; Xu, Y. Mining newsgroups using networks arising from social behavior. In Proceedings of the 12th International Conference on World Wide Web (WWW2003), Budapest, Hungary, 20 May 2003; pp. 529–535, doi:10.1145/775152.775227. [CrossRef]
14. Kwon, N.; Zhou, L.; Hovy, E.; Shulman, S.W. Identifying and classifying subjective claims. In Proceedings of the 8th Annual International Conference on Digital Government Research: Bridging Disciplines & Domains, New York, NY, USA, 20–23 May 2007; Digital Government Society of North America: Philadelphia, PA, USA, 2006; pp. 76–81.
15. Shulman, S.W. E-rulemaking: Issues in current research and practice. *Int. J. Public Adm.* **2015**, *28*, 621–641, doi:10.1081/PAD-200064221. [CrossRef]
16. Thomas, M.; Pang, B.; Lee, L. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP'06), Sydney, Australia, 22–23 July 2006; Association for Computational Linguistics: Stroudsburg, PA, USA, 2006; pp. 327–335.
17. Esuli, A.; Sebastiani, F. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, 22 May 2006.
18. Pang, B.; Lee, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd Meeting of the Association for Computational Linguistics, Barcelona, Spain, 21–26 July 2004; Association for Computational Linguistics: Stroudsburg, PA, USA, 2004; pp. 271–278, doi:10.3115/1218955.1218990. [CrossRef]
19. Yu, H.; Hatzivassiloglou, V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan, 11 July 2003; pp. 129–136.
20. Turney, P.D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; pp. 417–424, doi:10.3115/1073083.1073153. [CrossRef]
21. Pang, B.; Lee, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the 43rd Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA, 25–30 June 2005; Association for Computational Linguistics: Stroudsburg, PA, USA, 2005; pp. 115–124, doi:10.3115/1219840.1219855 . [CrossRef]
22. Wilson, T.; Wiebe, J.; Hwa, R. Just how mad are you? Finding strong and weak opinion clauses. In Proceedings of the 21st Conference of the American Association for Artificial Intelligence, Boston, MA, USA, 16–20 July 2006; AAAI Press: Palo Alto, CA, USA, 2004; pp. 761–769.
23. Baccianella, S.; Esuli, A.; Sebastiani, F. SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the International Conference on Language Resources and Evaluation, LREC, Valletta, Malta, 17–23 May 2010; pp. 2200–2204.
24. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135, doi:10.1561/1500000011. [CrossRef]
25. Wordnet. 2019. Available online: https://wordnet.princeton.edu/ (accessed on 28 June 2019).

26. Miller, G.A.; Beckwith, R.; Fellbaum, C.; Gross, D.; Miller, K.J. Introduction to WordNet: An On-line Lexical Database. *Int. J. Lexicogr.* **1990**, *3*, 235–244, doi:10.1093/ijl/3.4.235. [CrossRef]
27. Rauh, C. Validating a sentiment dictionary for German political language—A workbench note. *J. Inf. Technol. Polit.* **2018**, *15*, 319–343, doi:10.1080/19331681.2018.1485608. [CrossRef]
28. Young, L.; Soroka, S. Affective news: The automated coding of sentiment in political texts. *Polit. Commun.* **2012**, *29*, 205–231, doi:10.1080/10584609.2012.671234. [CrossRef]
29. Ceron, A.; Curini, L.; Iacus, S.M. iSA: A fast, scalable and accurate algorithm for sentiment analysis of social media content. *Inf. Sci.* **2016**, *367–368*, 105–124. [CrossRef]
30. Hopkins, D.; King, G. A method of automated nonparametric content analysis for social science. *Am. J. Polit. Sci.* **2010**, *54*, 229–247, doi:10.1111/j.1540-5907.2009.00428.x. [CrossRef]
31. Oliveira, D.J.S.; Bermejo, P.H.D.S.; dos Santos, P.A. Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. *J. Inf. Technol. Polit.* **2017**, *14*, 34–45, doi:10.1080/19331681.2016.1214094. [CrossRef]
32. Van Atteveldt, W.; Kleinnijenhuis, J.; Ruigrok, N.; Schlobach, S. Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *J. Inf. Technol. Polit.* **2008**, *5*, 73–94, doi:10.1080/19331680802154145. [CrossRef]
33. Klebanov, B.B.; Diermeier, D.; Beigman, E. Lexical cohesion analysis of political speech. *Polit. Anal.* **2008**, *16*, 447–463. [CrossRef]
34. Acharya, A.; Crawford, N.; Maduabum, M. *A Nation Divided: Classifying Presidential Speeches*; Stanford Univesity: Stanford, CA, USA, 2016.
35. Lakoff, G. *Moral Politics: How Liberals and Conservatives Think*, 2nd ed.; The University of Chicago Press: Chicago, IL, USA, 2002; doi:10.7208/chicago/9780226471006.001.0001. [CrossRef]
36. Lakoff, G.; Johnson, M. *Metaphors We Live By*; The Chicago University Press: Chicago, IL, USA, 1980.
37. Miner, G.; Elder, J.; Fast, A.; Hill, T.; Nisbet, R.; Delen, D. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data*; Academic Press: Cambridge, MA, USA, 2012.
38. Anurag, S.; Chatterjee, S.; Das, W.; Datta, D. Text Classification using Support Vector Machine. *Int. J. Eng. Sci. Invent.* **2015**, *4*, 33–37.
39. Lu, Y.; Wang, H.; Zhai, C.; Roth, D. Unsupervised discovery of opposing opinion networks from forum discussions. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, HI, USA, 2 November 2012; pp. 1642–1646.
40. Kennedy, A.; Inkpen, D. Sentiment classification of movie reviews using contextual valence shifters. *Comput. Intell.* **2006**, *22*, 110–125. [CrossRef]
41. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **2016**, *57*, 117–126. [CrossRef]
42. Joachims, T. Text categorization with Support Vector Machines: Learning with many relevant features. In *Machine Learning: ECML-98*; Nédellec, C., Rouveirol, C., Eds.; Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence); Springer: Berlin/Heidelberg, Germany, 1998; Volume 1398, pp. 137–142, doi:10.1007/BFb0026683. [CrossRef]
43. Sardianos, C.; Katakis, I.M.; Petasis, G.; Karkaletsis, V. Argument extraction from news. In Proceedings of the 2nd Workshop on Argumentation Mining, Denver, CO, USA, 4 June 2015; pp. 56–66. [CrossRef]
44. Florou, E.; Konstantopoulos, S.; Koukourikos, A.; Karampiperis, P. Argument extraction for supporting public policy formulation. In Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Sofia, Bulgaria, 8 August 2013; pp. 49–54.
45. Goudas, T.; Louizos, C.; Petasis, G.; Karkaletsis, V. Argument extraction from news, blogs, and social media. *Int. J. Artif. Intell. Tools* **2015**, *24*, 287–299, doi:10.1142/S0218213015400242. [CrossRef]
46. Lippi, M.; Torroni, P. Argument Mining from Speech: Detecting Claims in Political Debates. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12 February 2016; pp. 2979–2985; doi:10.5555/3016100.3016319. [CrossRef]
47. Sebastiani, F. Machine learning in automated text categorization. *ACM Comput. Surv.* **2002**, *34*, 1–47. [CrossRef]
48. Soumya, G.K.; Shibily, J. Text classification by augmenting Bag of Words (BOW) representation with co-occurrence feature. *OSR J. Comput. Eng.* **2014**, *16*, 34–38, doi:10.9790/0661-16153438. [CrossRef]

49. Giannakopoulos, G.; Mavridi, P.; Paliouras, G.; Papadakis, G.; Tserpes, K. Representation models for text classification: A comparative analysis over three web document types. In Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, Craiova, Romania, 13 June 2012; pp. 1–12; doi:10.1145/2254129.2254148. [CrossRef]
50. Gomaa, W.H.; Fahmy, A.A. A survey of text similarity approaches. *Int. J. Comput. Appl.* **2013**, *68*, 13–18.
51. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297, doi:10.1023/A:1022627411411. [CrossRef]
52. Vinodhini, G.; Chrasekaran, R.M. Sentiment Analysis and Opinion Mining: A Survey. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2012**, *2*, 282–292.
53. Berger, A.L.; Brown, P.F.; Della Pietra, S.A.; Della Pietra, V.J.; Gillett, J.R.; Lafferty, J.D.; Mercer, R.L.; Printz, H.; Ureš, L. The Candide system for machine translation. In *HLT '94 Proceedings of the Workshop on Human Language Technology*; Association for Computational Linguistics: Stroudsburg, PA, USA, 1994; pp. 157–162, doi:10.3115/1075812.1075844. [CrossRef]
54. Diermeier, D.; Godbout, J.-F.; Yu, B.; Kaufmann, S. Language and ideology in Congress. In Proceedings of the Annual Meeting of the Midwest Political Science Association (MPSA'07), Chicago, IL, USA, 4 April 2007.
55. Evans, M.; Wayne, M.; Cates, C.L.; Lin, J. Recounting the court? Toward a text-centered computational approach to understanding they dynamics of the judicial system. In Proceedings of the Annual Meeting of the Midwest Political Science Association, Chicago, IL, USA, 7 April 2005.
56. Laver, M.; Benoit, K.; Garry, J. Extracting policy positions from political texts using words as data. *Am. Polit. Sci. Rev.* **2003**, *97*, 311–337, doi:10.1017/S0003055403000698. [CrossRef]
57. Piryani, R.; Madhavi, D.; Singh, V.K. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Inf. Process. Manag.* **2017**, *53*, 122–150, doi:10.1016/j.ipm.2016.07.001 [CrossRef]
58. Riloff, E.; Wiebe, J. Learning extraction patterns for subjective expressions. In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-2003), Sapporo, Japan, 11–12 July 2003; Association for Computational Linguistics: Stroudsburg, PA, USA, 2003; pp. 105–112, doi:10.3115/1119355.1119369. [CrossRef]
59. Riloff, E.; Wiebe, J. Exploiting subjectivity classification to improve information extraction. In Proceedings of the 20th National Conference on Artificial Intelligence, Pittsburgh, PA, USA, 9–13 July 2005; AAAI Press: Palo Alto, CA, USA, 2005; Volume 3, pp. 1106–1111.
60. Lafferty, J.; McCallum, A.; Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, Williams College, MA, USA, 28 June–1 July 2001; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2001; pp. 282–289.
61. Riloff, E. An empirical study of automated dictionary construction for information extraction in three domains. *Artif. Intell.* **1996**, *85*, 101–134, doi:10.1016/0004-3702(95)00123-9 [CrossRef]
62. Choi, Y.; Cardie, C.; Riloff, E.; Patwardhan, S. Identifying sources of opinions with conditional random fields and extraction patterns. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Sydney, Australia, 22–23 July 2006; Association for Computational Linguistics: Stroudsburg, PA, USA, 2006; pp. 355–362, doi:10.3115/1220575.1220620. [CrossRef]
63. Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; Association for Computational Linguistics: Stroudsburg, PA, USA, 2005; pp. 347–354, doi:10.3115/1220575.1220619. [CrossRef]
64. Chesley, P.; Vincent, B.; Xu, L.; Srihari, R.K. Using verbs and adjectives to automatically classify blog sentiment. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs (2006)*; AAAI: Menlo Park, CA, USA, 2006.
65. Choi, Y.; Cardie, C. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–7 August 2009; Association for Computational Linguistics: Stroudsburg, PA, USA, 2009; Volume 2, pp. 590–598.

66. Jiang, L.; Yu, M.; Zhou, M.; Liu, X.; Zhao, T. Target-dependent twitter sentiment classification. In Proceedings of the 49th, Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; Volume1, pp. 151–160.
67. Tan, L.K.-W.; Na, J.-C.; Theng, Y.-L.; Chang, K. Sentence-Level Sentiment Polarity Classification Using a Linguistic Approach. In *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*; Xing, C., Crestani, F., Rauber, A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2011; Volume 7008, pp. 77–87, doi:10.1007/978-3-642-24826-9_13. [CrossRef]
68. Fang, X.; Zhan, J. Sentiment analysis using product review data. *J. Bigdata* **2015**, *2*, 5, doi:10.1186/s40537-015-0015-2. [CrossRef]
69. Nockleby, J.T. Hate Speech. In *Encyclopedia of the American Constitution*, 2nd ed.; Levy, L.W., Karst, K.L., Winkler, A., Eds.; Macmillan: New York, NY, USA, 2000; pp. 1277–1279.
70. Warner, W.; Hirschberg, J. Detecting Hate Speech on the World Wide Web. In Proceedings of the 2012 Workshop on Language in Social Media (LSM 2012), Montréal, QC, Canada, 7 June 2012; Association for Computational Linguistics: Stroudsburg, PA, USA, 2012; pp. 19–26.
71. Fiscus, J.G.; Ajot, J.; Garofolo, J.S. The Rich Transcription 2007 Meeting Recognition Evaluation. In *Multimodal Technologies for Perception of Humans. RT 2007, CLEAR 2007. Lecture Notes in Computer Science*; Stiefelhagen, R., Bowers, R., Fiscus, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; Volume 4625, pp. 373–389, doi:10.1007/978-3-540-68585-2_36. [CrossRef]
72. Camelin, N.; Béchet, F.; Damnati, G.; De Mori, R. Speech Mining in Noisy Audio Message Corpus. In Proceedings of the Interspeech 2007, Antwerp, Belgium, 27–31 August 2007; pp. 2401–2404. Available online: https://www.semanticscholar.org/paper/Speech-mining-in-noisy-audio-message-corpus-Camelin-Béchet/9d59c1f2d228fce67c5c6fac7f04cc1a2b29b532 (accessed on 15 January 2020).
73. Hookway, N. Entering the blogosphere: Some strategies for using blogs in social research. *Qual. Res.* **2008**, *8*, 91–113, doi:10.1177/1468794107085298. [CrossRef]
74. Thompson, C. The Early Years. *New York Magazine*, 10 February 2006, p. 1.
75. Tsai, F.S.; Chen, Y.; Chan, K.L. Probabilistic Techniques for Corporate Blog Mining. In *PAKDD 2007: Emerging Technologies in Knowledge Discovery and Data Mining*; Washio, T., Zhou, Z.-H., Huang, J.Z., Hu, X., Li, J., Xie, C., He, J., Zou, D., Li, K.-C., Freire, M.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 35–44, doi:10.1007/978-3-540-77018-3_5. [CrossRef]
76. Webb, L.M.; Wang, Y. Techniques for analyzing blogs and micro-blogs. In *Advancing Research Methods with New Technologies*; Sappleton, N., Ed.; IGI Global: Hershey, PA, USA, 2013; pp. 206–227, doi:10.4018/978-1-4666-3918-8.ch012. [CrossRef]
77. Tsai, F.S. Dimensionality reduction techniques for blog visualization. *Expert Syst. Appl.* **2011**, *38*, 2766–2773, doi:10.1016/j.eswa.2010.08.067. [CrossRef]
78. Tsai, F.S. A tag-topic model for blog mining. *Expert Syst. Appl.* **2011**, 38, 5330–5335, doi:10.1016/j.eswa.2010.10.025. [CrossRef]
79. Zafarani, R.; Abbasi, M.; Liu, H. *Social Media Mining: An Introduction*; Cambridge University Press: New York, NY, USA, 2014; doi:10.1017/CBO9781139088510. [CrossRef]
80. Berendt, B. Text mining for news and blogs analysis. In *Encyclopedia of Machine Learning and Data Mining*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2017; pp. 1247–1255, doi:10.1007/978-1-4899-7687-1. [CrossRef]
81. Barbier, G.; Liu, H. Data Mining in social media. In *Social Network Data Analytics*; Aggarwal, C.C., Ed.; Springer: Boston, MA, USA, 2011; pp. 327–352, doi:10.1007/978-1-4419-8462-3_12. [CrossRef]
82. Kumar, S.; Zafarani, R.; Abbasi, M.; Barbier, G.; Liu, H. Convergence of influential bloggers for topic discovery in the blogosphere. In *Advances in Social Computing. SBP 2010. Lecture Notes in Computer Science*; Chai, S.K., Salerno, J., Mabry, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6007, pp. 406–412, doi:10.1007/978-3-642-12079-4_51. [CrossRef]
83. Leban, G.; Fortuna, B.; Brank, J.; Grobelnik, M. Event registry: Learning about world events from news. In *WWW '14 Companion Proceedings of the 23rd International Conference on World Wide Web*; ACM: New York, NY, USA, 2014; pp. 107–110, doi:10.1145/2567948.2577024. [CrossRef]

84. Tsai, F.S.; Chan, K.L. Dimensionality reduction techniques for data exploration. In Proceedings of the 2007 6th International Conference on Information, Communications and Signal Processing, Singapore, 10–13 December 2007; pp. 1568–1572; doi:10.1109/ICICS.2007.4449863. [CrossRef]
85. Tsai, F.S.; Chan, K.L. Detecting Cyber Security Threats in Weblogs using Probabilistic Models. In *PAISI 2007: Intelligence and Security Informatics*; Yang, C.C., Zeng, D., Chau, M., Chang, K., Yang, Q., Cheng, X., Wang, J., Wang, F.-Y., Chen, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4430, pp. 46–57, doi:10.1007/978-3-540-71549-8_4. [CrossRef]
86. Liang, H.; Tsai, F.S.; Kdwee, A.T. Detecting novel business blogs. In Proceedings of the 7th International Conference on Information, Communications and Signal Processing, Macau, China, 8–10 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 651–655, doi:10.1109/ICICS.2009.5397541. [CrossRef]
87. Tsai, F.S. A data-centric approach to feed search in blogs. *Int. J. Web Eng. Technol.* **2012**, *7*, 228–249, doi:10.1504/ijwet.2012.048519. [CrossRef]
88. Tsai, F.S. Blogger-Link-Topic Model for Blog Mining. In *New Frontiers in Applied Data Mining. PAKDD 2011. Lecture Notes in Computer Science*; Cao, L., Huang, J.Z., Bailey, J., Koh, Y.S., Luo, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 28–39, doi:10.1007/978-3-642-28320-8_3. [CrossRef]
89. Tsai, F.S. Dimensionality reduction framework for blog mining and visualisation. *Int. J. Data Mining Model. Manag.* **2012**, *4*, 267–285, doi:10.1504/ijdmmm.2012.048108. [CrossRef]
90. Seep, K.S.; Patil, N. A Multidimensional Approach to Blog Mining. In *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*; Sa, P., Sahoo, M., Murugappan, M., Wu, Y., Majhi, B., Eds.; Springer: Singapore, 2018; pp. 51–58, doi:10.1007/978-981-10-3376-6_6. [CrossRef]
91. Tsirakis, N.; Poulopoulos, V.; Tsantilas, P.; Varlamis, I. Large scale opinion mining for social, news and blog data. *J. Syst. Softw.* **2017**, *127*, 237–248, doi:10.1016/j.jss.2016.06.012. [CrossRef]
92. Hussein, D.M.E.-D.M. A survey on sentiment analysis challenges. *J. King Saud Univ. Eng. Sci.* **2018**, *30*, 330–338, doi:10.1016/j.jksues.2016.04.002. [CrossRef]
93. Chen, M.-Y.; Chen, T.-H. Modeling public mood and emotion: Blog and news sentiment and socio-economic phenomena. *Future Gener. Comput. Syst.* **2019**, *96*, 692–699, doi:10.1016/j.future.2017.10.028. [CrossRef]
94. Tsai, F.S.; Chan, K.L. Blog Data Mining for Cyber Security Threats. In *Data Mining for Business Applications*; Cao, L., Yu, P.S., Zhang, C., Zhang, H., Eds.; Springer: Boston, MA, USA, 2009; pp. 169–182, doi:10.1007/978-0-387-79420-4_12. [CrossRef]
95. Lee, K.-C.; Hsieh, C.-H.; Wei, L.-J.; Mao, C.-H.; Dai, J.-H.; Kuang, Y.-T. Sec-Buzzer: Cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation. *Soft Comput.* **2017**, *21*, 2883–2896, doi:10.1007/s00500-016-2265-0. [CrossRef]
96. Valsamidis, S.; Theodosiou, T.; Kazanidis, I.; Nikolaidis, M. A Framework for opinion mining in blogs for agriculture. *Procedia Technol.* **2013**, *8*, 264–274. [CrossRef]
97. Kim, L.; Ju, J. Can media forecast technological progress? A text-mining approach to the on-line newspaper and blog's representation of prospective industrial technologies. *Inf. Process. Manag.* **2019**, *56*, 1506–1525, doi:10.1016/j.ipm.2018.10.017. [CrossRef]
98. Beheshti-Kashi, S.; Lùtjen, M.; Thoben, K.-D. Social media analytics for decision support in fashion buying processes. In *Artificial Intelligence for Fashion Industry in the Big Data Era, Springer Series in Fashion Business*; Thomassey, S., Zeng, X., Eds.; Springer: Singapore, 2018; pp. 71–93, doi:10.1007/978-981-13-0080-6_5. [CrossRef]
99. Bhadoria, R.S.; Dixit, M.; Bansal, R.; Chauhan, A.S. Detecting and searching system for event on internet blog data using cluster mining algorithm. In Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012), Visakhapatnam, India, 5–7 January 2012; Satapathy, S.C., Avadhani, P.S., Abraham, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 83–91, doi:10.1007/978-3-642-27443-5_10. [CrossRef]
100. Yuan, H.; Xu, H.; Qian, Y.; Li, Y. Make your travel smarter: Summarizing urban tourism information from massive blog data. *Int. J. Inf. Manag.* **2016**, *36*, 1306–1319, doi:10.1016/j.ijinfomgt.2016.02.009. [CrossRef]
101. Xu, H.; Yuan, H.; Ma, B.; Qian, Y. Where to go and what to play: Towards summarizing popular information from massive tourism blogs. *J. Inf. Sci.* **2019**, *41*, 830–854, doi:10.1177/0165551515603323. [CrossRef]
102. Evans, D.K.; Klavans, J.L.; McKeown, K.R. Columbia newsblaster: Multilingual news summarization on the web. In Proceedings of the Demonstration Papers at HLT-NAACL, Boston, MA, USA 2–7 May 2004. Available online: https://www.aclweb.org/anthology/N04-3001. (accessed on 15 January 2020).

103. Li, Z.; Tang, J.; Wang, X.; Liu, J.; Lu, H. Multimedia news summarization in search. *ACM Trans. Intell. Syst. Technol.* **2016**, *7*, 33. [CrossRef]
104. Kouris, P.; Alex, ridis, G.; Stafylopatis, A. Abstractive text summarization based on deep learning and semantic content generalization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July 2019; pp. 5082–5092.
105. Chen, Y.; Conroy, N.J.; Rubin, V.L. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, Seattle, WA, USA, 1 August 2015; pp. 15–19.
106. The Radicati Group, Inc. Email Statistics Report, 2019–2023–Executive Summary. February 2019. Available online: https://www.radicati.com/wp/wp-content/uploads/2018/12/Email-Statistics-Report-2019-2023-Executive-Summary.pdf (accessed on 1 January 2020).
107. Palmer, D.D. Text preprocessing. In *Handbook of Natural Language Processing*, 2nd ed.; Indurkhya, N., Damerau, F.J., Eds.; Chapman & Hall/CRC: London, UK, 2010; pp. 9–30.
108. Katakis, I.; Tsoumakas, G.; Vlahavas, I. E-mail mining: Emerging techniques for E-Mail management. In *Web Data Management Practices: Emerging Techniques and Technologies*; Vakali, A., Pallis, G., Eds.; IGI Global: Hershey, PA, USA, 2007; pp. 220–243, doi:10.4018/978-1-59904-228-2.ch010. [CrossRef]
109. Laclavík, M.; Dlugolinský, Š.; Šeleng, M.; Kvassay, M.; Gatial, E.; Balogh, Z.; Hluchý, L. Email analysis and information extraction for enterprise benefit. *Comput. Inform.* **2011**, *30*, 57–87.
110. Chen, F.; Deng, P.; Wan, J.; Zhang, D.; Vasilakos, A.V.; Rong, X. Data mining for the internet of things: literature review and challenges. *Int. J. Distrib. Sens. Netw.* **2015**, 431047, doi:10.1155/2015/431047 [CrossRef]
111. Wani, M.A.; Jabin, S. Big Data: Issues, challenges, and techniques in business intelligence. In *Big Data Analytics. Advances in Intelligent Systems and Computing*; Aggarwal, V., Bhatnagar, V., Mishra, D., Eds.; Springer: Singapore, 2018; pp. 613–628, doi:10.1007/978-981-10-6620-7_59. [CrossRef]
112. Klimt, B.; Yang, Y. Introducing the Enron corpus. In Proceedings of the CEAS 2004—First Conference on Email and Anti-Spam, Mountain View, CA, USA, 30–31 July 2004.
113. Minkov, E.; Wang, R.C.; Cohen, W.W. Extracting personal names from emails: Applying named entity recognition to informal text. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; Association for Computational Linguistics: Stroudsburg, PA, USA, 2005; pp. 443–450, doi:10.3115/1220575.1220631. [CrossRef]
114. Androutsopoulos, I.; Koutsias, J.; Chrinos, K.V.; Paliouras, G.; Spyropoulos, C. An evaluation of naive Bayesian anti-spam filtering. In Proceedings of the 1th European Conference on Machine Learning in the New Information Age, Barcelona, Spain, 2 June 2000; pp. 9–17.
115. Weerkamp, W.; Balog, K.; De Rijke, M. Using contextual information to improve search in email archives. In Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, Toulouse, France, 6–9 April 2009; Springer: Berlin/Heidelberg, Germany, 2009; pp. 400–411, doi:10.1007/978-3-642-00958-7_36 [CrossRef]
116. Tang, G.; Pei, J.; Luk, W.S. Email mining: Tasks, common techniques, and tools. *Knowl. Inf. Syst.* **2014**, *41*, 1–31. [CrossRef]
117. Mujtaba, G.; Shuib, L.; Raj, R.G.; Majeed, N.; Al-Garadi, M.A. Email classification research trends: review and open issues. *IEEE Access* **2017**, *5*, 9044–9064, doi:10.1109/access.2017.2702187. [CrossRef]
118. Hangal, S.; Lam, M.S.; Heer, J. MUSE: Reviving memories using email archives. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, 16–19 October 2011; ACM: New York, NY, USA, 2011; pp. 75–84, doi:10.1145/2047196.2047206. [CrossRef]
119. Liu, B. *Sentiment Analysis and Opinion Mining*; Morgan & Claypool Publishers: Williston, VT, USA, 2012, doi:10.2200/s00416ed1v01y201204hlt016. [CrossRef]
120. Liu, S.; Lee, I. A Hybrid Sentiment Analysis Framework for Large Email Data. In Proceedings of the 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Taipei, Taiwan, 24–27 November 2015; IEEE: Piscataway, NJ, USA, 2015, doi:10.1109/iske.2015.91. [CrossRef]
121. Liu, S.; Lee, I. Discovering sentiment sequence within email data through trajectory representation. *Expert Syst. Appl.* **2018**, *99*, 1–11, doi:10.1016/j.eswa.2018.01.026. [CrossRef]
122. Wimmer, B. *Business Espionage: Risk, Threats, and Countermeasures*; Butterworth-Heinemann: Oxford, UK, 2015; doi:10.1016/C2013-0-09869-6. [CrossRef]

123. Chi, H.; Scarllet, C.; Prodanoff, Z.G.; Hubbard, D. Determining predisposition to insider threat activities by using text analysis. In *Future Technologies Conference (FTC)*; IEEE: Piscataway, NJ, USA, 2016; pp. 985–990, doi:10.1109/ftc.2016.7821723. [CrossRef]
124. Soh, C.; Yu, S.; Narayanan, A.; Duraisamy, S.; Chen, L. Employee profiling via aspect-based sentiment and network for insider threats detection. *Expert Syst. Appl.* **2019**, 351–361, doi:10.1016/j.eswa.2019.05.043. [CrossRef]
125. Cisco Talos Intelligence Group Report. 2019. Available online: https://www.talosintelligence.com/ (accessed on 1 January 2020)
126. Osterman Research, Inc. *Techniques for Dealing with Ransomware, Business Email Compromise and Spearphishing, An Osterman Research White Paper*; Osterman Research, Inc.: Washington, DC, USA, 2017.
127. Tretyakov, K. Machine Learning Techniques in Spam Filtering. In *Data Mining Problem-Oriented Seminar*; MTAT: Beauvallon, France, 2004; pp. 60–79. Available online: https://courses.cs.ut.ee/2004/dm-seminar-spring/uploads/Main/P06.pdf (accessed on 1 January 2020).
128. Bhowmick, A.; Hazarika, S.M. Machine learning for E-Mail spam filtering: review, techniques and trends. *arXiv* **2016**, arXiv:1606.01042v1.
129. Dada, E.G.; Bassi, J.S.; Chiroma, H.; Abdulhamid, S.M.; Adetunmbi, A.O.; Ajibuwa, O.E. Machine learning for email spam filtering: Review, approaches and open research problems. *Heliyon* **2019**, *5*, e01802, doi:10.1016/j.heliyon.2019.e01802. [CrossRef]
130. Bahgat, E.M.; Rady, S.; Gad, W.; Moawad, I.F. Efficient email classification approach based on semantic methods. *Ain Shams Eng. J.* **2018**, *9*, 3259–3269, doi:10.1016/j.asej.2018.06.001. [CrossRef]
131. Almomani, A.; Wan, T.C.; Manasrah, A.; Altaher, A.; Baklizi, M.; Ramadass, S. An enhanced online phishing e-mail detection framework based on evolving connectionist system. *Int. J. Innov. Comput. Inf. Control* **2013**, *9*, 169–175.
132. Chowdhury, M.U.; Abawajy, J.H.; Kelarev, A.V.; Hochin, T. Multilayer hybrid strategy for phishing email zero-day filtering. *Concurr. Comput. Pract. Exp.* **2016**, *29*, e3929, doi:10.1002/cpe.3929 [CrossRef]
133. Smadi, S.; Aslam, N.; Zhang, L. Detection of online phishing email using dynamic evolving neural network based on reinforcement learning. *Decis. Support Syst.* **2018**, *107*, 88–102, doi:10.1016/j.dss.2018.01.001. [CrossRef]
134. Gök, A.; Waterworth, A.; Shapira, P. Use of web mining in studying innovation. *Scientometrics* **2015**, *102*, 653–671, doi:10.1007/s11192-014-1434-0. [CrossRef]
135. Waldherr, A.; Maier, D.; Miltner, P.; Günther, E. B Big Data, Big Noise: The Challenge of Finding Issue Networks on the Web. *Soc. Sci. Comput. Rev.* **2017**, *35*, 427–443, doi:10.1177/0894439316643050 [CrossRef]
136. Etzioni, O. The world wide web: Quagmire or gold mine. *Commun. ACM* **1996**, *39*, 65–68, doi:10.1145/240455.240473. [CrossRef]
137. Cooley, R.; Mobasher, B.; Srivastava, J. Data preparation for mining World Wide Web browsing patterns. *Knowl. Inf. Syst.* **1999**, *1*, 5–32, doi:10.1007/BF03325089. [CrossRef]
138. Markov, Z.; Larose, D.T. *Data Mining the Web: Uncovering Patterns in Web Content, Structure and Usage*; Wiley-Interscience: Hoboken, NJ, USA, 2007.
139. Velásquez, J.D. Web mining and privacy concerns: Some important legal issues to be consider before applying any data and information extraction technique in web-based environments. *Expert Syst. Appl.* **2013**, *40*, 5228–5239, doi:10.1016/j.eswa.2013.03.008. [CrossRef]
140. Borges, J.; Levene, M. Data mining of user navigation patterns. In *Web Usage Analysis and User Profiling. WebKDD 1999. Lecture Notes in Computer Science*; Masand, B., Spiliopoulou, M., Eds.; Springer: Berlin/Heidelberg, Germany, 1999; pp. 92–112, doi:10.1007/3-540-44934-5_6. [CrossRef]
141. Madria, S.K.; Bhowmick, S.S.; Ng, W.K.; Lim, E.P. Research Issues in Web Data Mining. In *DataWarehousing and Knowledge Discovery. DaWaK 1999. Lecture Notes in Computer Science*; Mohania, M., Tjoa, A.M., Eds.; Springer: Berlin/Heidelberg, Germany, 1999; pp. 303–312, doi:10.1007/3-540-48298-9_32. [CrossRef]
142. Xu, G.; Zhang, Y.; Li, L. *Web Mining and Social Networking*; Springer: Boston, MA, USA, 2011, doi:10.1007/978-1-4419-7735-9. [CrossRef]
143. Kanathey, K.; Thakur, R.S.; Jaloree, S. Ranking of web pages using aggregation of page rank and hits algorithm. *Int. J. Adv. Stud. Comput. Sci. Eng.* **2018**, *7*, 17–22.
144. Facca, F.M.; Lanzi, P.L. Mining interesting knowledge from weblogs: A survey. *Data Knowl. Eng.* **2005**, *53*, 225–241, doi:10.1016/j.datak.2004.08.001. [CrossRef]

145. Srivastava, J.; Cooley, R.; Deshpe, M.; Tan, P.-N. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explor. Newsl.* **2000**, *1*, 12–23, doi:10.1145/846183.846188 [CrossRef]
146. Liu, H.; Keselj, V. Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users' future requests. *Data Knowl. Eng.* **2007**, *61*, 304–330, doi:10.1016/j.datak.2006.06.001. [CrossRef]
147. Kohli, S.; Gupta, A. Fuzzy information retrieval in WWW: A survey. *Int. J. Adv. Intell. Paradig.* **2014**, *6*, 272–311, doi:10.1504/IJAIP.2014.066984. [CrossRef]
148. Gupta, A.; Kohli, S. FORA: An OWO based framework for finding Outliers in Web Usage Mining. *Inf. Fusion* **2019**, *48*, 27–38, doi:10.1016/j.inffus.2018.08.003. [CrossRef]
149. Chola, V.; A Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 15, doi:10.1145/1541880.1541882. [CrossRef]
150. Gupta, A.; Kohli, S. An analytical study of ordered weighted geometric averaging operator on Web data set as a MCDM problem. In Proceedings of the Fourth International Conference on Soft Computing for Problem Solving, Assam, India, 23 December 2014; Das, K., Deep, K., Pant, M., Bansal, J., Nagar, A., Eds.; Springer: New Delhi, India, 2014; pp. 585–597, doi:10.1007/978-81-322-2217-0_47. [CrossRef]
151. Gupta, A.; Kohli, S. OWA operator-based hybrid framework for outlier reduction in web mining. *Int. J. Intell. Syst.* **2016**, *31*, 947–962, doi:10.1002/int.21810. [CrossRef]
152. Iglesias, J.A.; Tiemblo, A.; Ledezma, A.; Sanchis, A. Web news mining in an evolving framework. *Inf. Fusion* **2016**, *28*, 90–98, doi:10.1016/j.inffus.2015.07.004. [CrossRef]
153. Za'in, C.; Pratama, M.; Lughofer, E.; Anavatti, S.G. Evolving type-2 web news mining. *Appl. Soft Comput.* **2017**, *54*, 200–220, doi:10.1016/j.asoc.2016.11.034. [CrossRef]
154. Kosala, R.; Blockeel, H. Web mining research: A survey. *ACM SIGKDD Explor. Newsl.* **2000**, *2*, 1–15, doi:10.1145/360402.360406. [CrossRef]
155. Dias, J.P.; Ferreira, H.S. Automating the extraction of static content and dynamic behaviour from e-commerce websites. *Procedia Comput. Sci.* **2017**, *109*, 297–304, doi:10.1016/j.procs.2017.05.355. [CrossRef]
156. Zhou, J.; Cheng, C.; Kang, L.; Sun, R. Integration and Analysis of Agricultural Market Information Based on Web Mining. *IFAC-PapersOnLine* **2018**, *51*, 778–783, doi:10.1016/j.ifacol.2018.08.101. [CrossRef]
157. Symantec Corporation Inc. Internet Security Threat Report. 2019. Available online: https://resource.elq.symantec.com/LP=6819?CID=70138000001QvI4AAK (accessed on 1 January 2020).
158. Mohammad, R.M.; Thabtah, F.; McCluskey, L. Tutorial and critical analysis of phishing websites methods. *Comput. Sci. Rev.* **2015**, *17*, 1–24, doi:10.1016/j.cosrev.2015.04.001. [CrossRef]
159. Yi, P.; Guan, Y.; Zou, F.; Yao, Y.; Wang, W.; Zhu, T. Web Phishing Detection Using a Deep Learning Framework. *Wirel. Commun. Mob. Comput.* **2018**, 1–9, doi:10.1155/2018/4678746. [CrossRef]
160. Román, P.E.; Dell, R.F.; Velásquez, J.D.; Loyola, P.S. Identifying User Sessions from Web Server Logs with Integer Programming. *Intell. Data Anal.* **2014**, *18*, 43–61, doi:10.3233/IDA-130627. [CrossRef]
161. Apaolaza, A.; Vigo, M. Assisted pattern mining for discovering interactive behaviors on the web. *Int. J. Hum.-Comput. Stud.* **2019**, *130*, 196–208, doi:10.1016/j.ijhcs.2019.06.012. [CrossRef]
162. Slanzi, G.; Pizarro, G.; Velásquez, J.D. Biometric information fusion for web user navigation and preferences analysis: An overview. *Inf. Fusion* **2017**, *38*, 12–21, doi:10.1016/j.inffus.2017.02.006. [CrossRef]
163. Öztürk, N.; Ayvaz, S. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telemat. Inf.* **2018**, *35*, 136–147, doi:10.1016/j.tele.2017.10.006. [CrossRef]
164. Irfan, R.; King, C.K.; Grages, D.; Ewen, S.; Khan, S.U.; Madani, S.A.; Kolodziej, J.; Wang, L.; Chen, D.; Rayes, A.; et al. A survey on text mining in social networks. *Knowl. Eng. Rev.* **2015**, *30*, 157–170, doi:10.1017/S0269888914000277. [CrossRef]
165. Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 17–23 May 2010; Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D., Eds.; European Language Resources Association (ELRA): Luxembourg, 2010; pp. 1320–1326.
166. Nisar, T.M.; Yeung, M. Twitter as a tool for forecasting stock market movements: A short-window event study. *J. Financ. Data Sci.* **2018**, *4*, 101–119, doi:10.1016/j.jfds.2017.11.002. [CrossRef]
167. Bollen, J.; Mao, H.; Zeng, X. Twitter mood predicts the stock market. *J. Comput. Sci.* **2011**, *2*, 1–8, doi:10.1016/j.jocs.2010.12.007. [CrossRef]

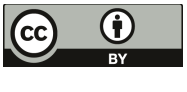
168. Ruiz, E.J.; Hristidis, V.; Castillo, C.; Gionis, A.; Jaimes, A. Correlating financial time series with micro-blogging activity. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, (WSDM'12), Seattle, WA, USA, 8–12 February 2012; ACM: New York, NY, USA, 2012; pp. 513–522, doi:10.1145/2124295.2124358. [CrossRef]
169. Hagenau, M.; Liebmann, M.; Neumann, D. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decis. Support Syst.* **2013**, *55*, 685–697, doi:10.1016/j.dss.2013.02.006. [CrossRef]
170. Zhang, L. *Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation*; The University of Texas: Austin, TX, USA, 2013.
171. Bing, L.; Chan, K.C.; Ou, C. Public sentiment analysis in Twitter data for prediction of a company's stock price movements. In Proceedings of the 2014 IEEE 11th International Conference on e-Business Engineering, Guangzhou, China, 5–7 November 2014; IEEE: Piscataway, NJ, USA, 2014, doi:10.1109/ICEBE.2014.47. [CrossRef]
172. Dickinson, B.; Hu, W. Sentiment analysis of investor opinions on twitter. *Soc. Netw.* **2015**, *4*, 62–71, doi:10.4236/sn.2015.43008. [CrossRef]
173. Das, S.; Behera, R.K.; Rath, S.K. Real-time sentiment analysis of Twitter streaming data for stock prediction. *Procedia Comput. Sci.* **2018**, *132*, 956–964, doi:10.1016/j.procs.2018.05.111. [CrossRef]
174. Alkubaisi, G.A.A.J.; Kamaruddin, S.S.; Husni, H. Stock market classification model using sentiment analysis on twitter based on hybrid naive bayes classifiers. *Comput. Inf. Sci.* **2018**, *11*, 52–64, doi:10.5539/cis.v11n1p52. [CrossRef]
175. Broadstock, D.C.; Zhang, D. Social-media and intraday stock returns: The pricing power of sentiment. *Financ. Res. Lett.* **2019**, 116–123, doi:10.1016/j.frl.2019.03.030. [CrossRef]
176. Alkhatib, M.; El Barachi, M.; Shaalan, K. An Arabic social media based framework for incidents and events monitoring in smart cities. *J. Clean. Prod.* **2019**, *220*, 771–785, doi:10.1016/j.jclepro.2019.02.063. [CrossRef]
177. Gupta, B.; Sharma, S.; Chennamaneni, A. Twitter Sentiment Analysis: An Examination of Cybersecurity Attitudes and Behavior. In Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics, Dublin, Ireland, 11 December 2016; p. 17.
178. Philer, K.; Zhong, Y. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *Int. J. Hosp. Manag.* **2016**, *55*, 16–24, doi:10.1016/j.ijhm.2016.02.001. [CrossRef]
179. Lee, N.Y.; Kim, Y.; Sang, Y. How do journalists leverage Twitter? Expressive and consumptive use of Twitter. *Soc. Sci. J.* **2017**, *54*, 139–147, doi:10.1016/j.soscij.2016.09.004. [CrossRef]
180. Crannell, W.C.; Clark, E.; Jones, C.; James, T.A.; Moore, J. A pattern-matched Twitter analysis of US cancer-patient sentiments. *J. Surg. Res.* **2016**, *206*, 536–542, doi:10.1016/j.jss.2016.06.050. [CrossRef]
181. Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F.; Narayanan, S. A system for real-time twitter sentiment analysis of 2012 US presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations, Jeju Island, Korea, 8–14 July 2012; Association for Computational Linguistics: Stroudsburg, PA, USA, 2012; pp. 115–120.
182. Greco, F.; Polli, A. Emotional text mining: Customer profiling in brand management. *Int. J. Inf. Manag.* **2019**, doi:10.1016/j.ijinfomgt.2019.04.007. [CrossRef]
183. Akundi, A.; Tseng, B.; Wu, J.; Smith, E.; Subbalakshmi, M.; Aguirre, F. Text mining to understand the influence of social media applications on smartphone supply chain. *Procedia Comput. Sci.* **2018**, *140*, 87–94, doi:10.1016/j.procs.2018.10.296. [CrossRef]
184. Mansour, S. Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. *Procedia Comput. Sci.* **2018**, *140*, 95–103, doi:10.1016/j.procs.2018.10.297. [CrossRef]
185. Reyes-Menendez, A.; Saura, J.R.; Alvarez-Alonso, C. Understanding #WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2537, doi:10.3390/ijerph15112537. [CrossRef]
186. Al-Daihani, S.M.; Abrahams, A. A text mining analysis of academic libraries' Tweets. *J. Acad. Librariansh.* **2016**, *42*, 135–143, doi:10.1016/j.acalib.2015.12.014. [CrossRef]
187. Center, P.R. *Social Media Fact Sheet*; Pew Research Center: Washington, DC, USA, 2017.
188. Kim, J.; Hastak, M. Social network analysis: Characteristics of online social networks after a disaster. *Int. J. Inf. Manag.* **2018**, *38*, 86–96, doi:10.1016/j.ijinfomgt.2017.08.003. [CrossRef]
189. He, W.; Zha, S.; Li, L. Social media competitive analysis and text mining: A case study in the pizza industry. *Int. J. Inf. Manag.* **2013**, *33*, 464–472, doi:10.1016/j.ijinfomgt.2013.01.001. [CrossRef]

190. Salloum, S.A.; Mhamdi, C.; Al-Emran, M.; Shaalan, K. Analysis and classification of Arabic newspapers' Facebook pages using text mining techniques. *Int. J. Inf. Technol. Lang. Stud.* **2017**, *1*, 8–17.
191. Al-Daihani, S.M.; Abrahams, A. Analysis of academic libraries' facebook posts: Text and data analytics. *J. Acad. Librariansh.* **2018**, *44*, 216–225, doi:10.1016/j.acalib.2018.02.004. [CrossRef]
192. Serna, A.; Gasparovic, S. Transport analysis approach based on big data and text mining analysis from social media. *Transp. Res. Procedia* **2018**, *33*, 291–298, doi:10.1016/j.trpro.2018.10.105. [CrossRef]
193. Sezgen, E.; Mason, K.J.; Mayer, R. Voice of airline passenger: A text mining approach to understand customer satisfaction. *J. Air Transp. Manag.* **2019**, *77*, 65–74, doi:10.1016/j.jairtraman.2019.04.001. [CrossRef]
194. Suresh, V.; Roohi, S.; Eirinaki, M. Aspect-based opinion mining and recommendation system for restaurant reviews. In Proceedings of the 8th ACM Conference on Recommender systems, Foster City, CA, USA, 1 October 2014; pp. 361–362, doi:10.1145/2645710.2645716. [CrossRef]
195. Saha, S.; Santra, A.K. Restaurant rating based on textual feedback. In Proceedings of the 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS), Vellore, India, 10–12 August 2017, doi:10.1109/ICMDCS.2017.8211542. [CrossRef]
196. Chen, M.-Y.; Liao, C.-H.; Hsieh, R.-P. Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach. *Comput. Hum. Behav.* **2019**, doi:10.1016/j.chb.2019.03.021 [CrossRef]
197. Liu, Y.; Qin, Z.; Li, P.; Wan, T. Stock volatility prediction using recurrent neural networks with sentiment analysis. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*; Benferhat, S., Tabia, K., Ali, M., Eds.; Springer: Cham, Switzerland, 2017; pp. 192–201, doi:10.1007/978-3-319-60042-0_22. [CrossRef]
198. Chen, W.; Yeo, C.K.; Lau, C.T.; Lee, B.S. Leveraging social media news to predict stock index movement using RNN-boost. *Data Knowl. Eng.* **2018**, *118*, 14–24, doi:10.1016/j.datak.2018.08.003 [CrossRef]
199. Liu, P.; Xia, X.; Li, A. Tweeting the financial market: Media effect in the era of Big Data. *Pac. Basin Financ. J.* **2018**, *51*, 267–290, doi:10.1016/j.pacfin.2018.07.007. [CrossRef]
200. Zhang, X.; Shi, J.; Wang, D.; Fang, B. Exploiting investors social network for stock prediction in China's market. *J. Comput. Sci.* **2018**, *28*, 294–303, doi:10.1016/j.jocs.2017.10.013. [CrossRef]
201. Pejic-Bach, M.; Bertoncel, T.; Meško, M.; Krstic, Ž. Text mining of industry 4.0 job advertisements. *Int. J. Inf. Manag.* **2019**, doi:10.1016/j.ijinfomgt.2019.07.014. [CrossRef]
202. Moro, S.; Cortez, P.; Rita, P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Syst. Appl.* **2015**, *42*, 1314–1324, doi:10.1016/j.eswa.2014.09.024. [CrossRef]
203. Amado, A.; Cortez, P.; Rita, P.; Moro, S. Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *Eur. Res. Manag. Bus. Econ.* **2018**, *24*, 1–7, doi:10.1016/j.iedeen.2017.06.002. [CrossRef]
204. Moro, S.; Pires, G.; Rita, P.; Cortez, P. A text mining and topic modelling perspective of ethnic marketing research. *J. Bus. Res.* **2019**, *103*, 275–285, doi:10.1016/j.jbusres.2019.01.053. [CrossRef]
205. Cortez, P.; Moro, S.; Rita, P.; King, D.; Hall, J. Insights from a text mining survey on Expert Systems research from 2000 to 2016. *Expert Syst.* **2018**, *35*, e12280, doi:10.1111/exsy.12280. [CrossRef]
206. Moro, S.; Rita, P. Brand strategies in social media in hospitality and tourism. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 343–364, doi:10.1108/IJCHM-07-2016-0340. [CrossRef]
207. Guerreiro, J.; Rita, P.; Trigueiros, D. A text mining-based review of cause-related marketing literature. *J. Bus. Ethics* **2016**, *139*, 111–128, doi:10.1007/s10551-015-2622-4. [CrossRef]
208. Loureiro, S.M.C.; Guerreiro, J.; Eloy, S.; Langaro, D.; Panchapakesan, P. Understanding the use of virtual reality in marketing: A text mining-based review. *J. Bus. Res.* **2019**, *100*, 514–530, doi:10.1016/j.jbusres.2018.10.055. [CrossRef]
209. Galati, F.; Bigliardi, B. Industry 4.0: Emerging themes and future research avenues using a text mining approach. *Comput. Ind.* **2019**, *109*, 100–113, doi:10.1016/j.compind.2019.04.018. [CrossRef]
210. Guan, J.; Manikas, A.S.; Boyd, L.H. The at 55: A content-driven review and analysis. *Int. J. Prod. Res.* **2017**, *57*, 4667–4675, doi:10.1080/00207543.2017.1296979. [CrossRef]
211. Demeter, K.; Szász, L.; Kö, A. A text mining based overview of inventory research in the ISIR special issues 1994-2016. *Int. J. Prod. Econ.* **2018**, *209*, 134–146, doi:10.1016/j.ijpe.2018.06.006. [CrossRef]
212. Grubert, E. Implicit prioritization in life cycle assessment: Text mining and detecting metapatterns in the literature. *Int. J. Life Cycle Assess.* **2016**, *22*, 148–158, doi:10.1007/s11367-016-1153-2. [CrossRef]

213. Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.C.; Coimbra, C.F.M. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Sol. Energy* **2018**, *168*, 60–101, doi:10.1016/j.solener.2017.11.023. [CrossRef]
214. Moro, S.; Rita, P.; Cortez, P. A text mining approach to analyzing Annals literature. *Ann. Tour. Res.* **2017**, *66*, 208–210, doi:10.1016/j.annals.2017.07.011 [CrossRef]
215. Contiero, B.; Cozzi, G.; Karpf, L.; Gottardo, F. Pain in Pig Production: Text Mining Analysis of the Scientific Literature. *J. Agric. Environ. Ethics* **2019**, *32*, 401–412, doi:10.1007/s10806-019-09781-4. [CrossRef]
216. Wang, S.-H.; Ding, Y.; Zhao, W.; Huang, Y.-H.; Perkins, R.; Zou, W.; Chen, J.J. Text mining for identifying topics in the literatures about adolescent substance use and depression. *BMC Public Health* **2016**, *16*, doi:10.1186/s12889-016-2932-1. [CrossRef]
217. Balan, P.F.; Gerits, A.; Vuffel, W. A practical application of text mining to literature on cognitive rehabilitation and enhancement through neurostimulation. *Front. Syst. Neurosci.* **2014**, *8*, 182, doi:10.3389/fnsys.2014.00182. [CrossRef]
218. Carvalho, A.S., Rodríguez, M.S. and Matthiesen, R. Review and literature mining on proteostasis factors and cancer. In *Proteostasis. Methods in Molecular Biology*; Matthiesen, R., Ed.; Humana Press: New York, NY, USA, 2016; pp. 71–84, doi:10.1007/978-1-4939-3756-1_2. [CrossRef]
219. Karami, A.; Ghasemi, M.; Sen, S.; Moraes, M.F.; Shah, V. Exploring diseases and syndromes in neurology case reports from 1955 to 2017 with text mining. *Comput. Biol. Med.* **2019**, *109*, 322–332, doi:10.1016/j.compbiomed.2019.04.008. [CrossRef] [PubMed]
220. Kayal, S.; Afzal, Z.; Tsatsaronis, G.; Doornenbal, M.; Katrenko, S.; Gregory, M. A framework to automatically extract funding information from text. In Proceedings of the International Conference on Machine Learning, Optimization, and Data Science, Volterra, Italy, 13 September 2018; pp. 317–328.
221. Yousif, A.; Niu, Z.; Nyamawe, A.S.; Hu, Y. Improving citation sentiment and purpose classification using hybrid deep neural network model. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, Cairo, Egypt, 26–28 October 2018; pp. 327–336.
222. Sag, M. The new legal landscape for text mining and machine learning. *J. Copyr. Soc. USA* **2019**, *66*, doi:10.2139/ssrn.3331606. [CrossRef]
223. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright in the Digital Single Market. Available online: https://eur-lex.europa.eu/eli/dir/2019/790/oj (accessed on 1 January 2020)
224. Sheeba, J.; Vivekanan, K. Improved keyword and keyphrase extraction from meeting transcripts. *Int. J. Comput. Appl.* **2012**, *52*, 11–15.
225. Liu, F.; Liu, F.; Liu, Y. A supervised framework for keyword extraction from meeting transcripts. *IEEE Trans. Audio Speech Lang. Process.* **2010**, *19*, 538–548, doi:10.1109/TASL.2010.2052119. [CrossRef]
226. Liu, F.; Pennell, D.; Liu, F.; Liu, Y. Unsupervised approaches for automatic keyword extraction using meeting transcripts. In *NAACL'09 Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2009; pp. 620–628.
227. Song, H.-J.; Go, J.; Park, S.-B.; Park, S.-Y. A just-in-time keyword extraction from meeting transcripts. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 9–14 June 2013; Association for Computational Linguistics: Atlanta, GA, USA, 2013; pp. 888–896.
228. Song, H.-J.; Go, J.; Park, S.-B.; Park, S.-Y.; Kim, K.Y. A just-in-time keyword extraction from meeting transcripts using temporal and participant information. *J. Intell. Inf. Syst.* **2017**, *48*, 117–140, doi:10.1007/s10844-015-0391-2. [CrossRef]
229. Xie, S.; Liu, Y. Improving supervised learning for meeting summarization using sampling and regression. *Comput. Speech Lang.* **2010**, *24*, 495–514, doi:10.1016/j.csl.2009.04.007. [CrossRef]
230. Sharp, B.; Chibelushi, C. Text segmentation of spoken meeting transcripts. *Int. J. Speech Technol.* **2008**, *11*, 157, doi:10.1007/s10772-009-9048-2. [CrossRef]
231. Amancio, D.R.; Altmann, E.G.; Oliveira, O.N., Jr.; Costa, L.F. Comparing intermittency and network measurements of words and their dependence on authorship. *New J. Phys.* **2011**, *13*, 123024. [CrossRef]
232. Amancio, D.R.; Oliveira, O.N., Jr.; Costa, L.F. Identification of literary movements using complex networks to represent texts. *New J. Phys.* **2012**, *14*, 043029. [CrossRef]

233. Amancio, D.R. A complex network approach to stylometry. *PLoS ONE* **2015**, *10*, e0136076, doi:10.1371/journal.pone.0136076. [CrossRef]
234. Wang, T.; Brede, M.; Ianni, A.; Mentzakis, E. Characterizing dynamic communication in online eating disorder communities: A multiplex network approach. *Appl. Netw. Sci.* **2019**, *4*, doi:10.1007/s41109-019-0125-4. [CrossRef]
235. Nuzzo, A.; Mulas, F.; Gabetta, M.; Arbustini, E.; Zupan, B.; Larizza, C.; Bellazzi, R. Text mining approaches for automated literature knowledge extraction and representation. *Stud. Health Technol. Inform.* **2010**, *160*, 954–958, doi:10.3233/978-1-60750-588-4-954. [CrossRef] [PubMed]
236. Gentzkow, M.; Kelly, B.T.; Taddy, M. Text As Data. *NBER Work. Pap.* **2017**, doi:10.2139/ssrn.2934001. [CrossRef]
237. Lau, R.; Xia, Y. Latent text mining for cybercrime forensics. *Int. J. Future Comput. Commun.* **2013**, *2*, 368–371, doi:10.7763/ijfcc.2013.v2.187. [CrossRef]
238. Suh-Lee, C.; Ju-Yeon, J.; Yoohwan, K. Text mining for security threat detection discovering hidden information in unstructured log messages. In Proceedings of the 2016 IEEE Conference on Communications and Network Security (CNS), Philadelphia, PA, USA, 17–19 October 2016; IEEE: Piscataway, NJ, USA, 2016, doi:10.1109/CNS.2016.7860492. [CrossRef]
239. Noel, S. Text Mining for Modeling Cyberattacks. In *Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications*; Venkat, N., Gudivada, C.R., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; Chapter 14, pp. 463–515, doi:10.1016/bs.host.2018.06.001. [CrossRef]
240. Dong, F.; Yuan, S.; Ou, H.; Liu, L. New Cyber Threat Discovery from Darknet Marketplaces. In Proceedings of the IEEE Conference on Big Data and Analytics (ICBDA), Shanghai, China, 21–22 November 2018; IEEE: Piscataway, NJ, USA, 2018, doi:10.1109/icbdaa.2018.8629658. [CrossRef]
241. Kaplan, J.; Sharma, S.; Weinberg, A. Meeting the Cybersecurity Challenge. Available online: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/meeting-the-cybersecurity-challenge (accessed on 1 January 2020).
242. Aiken, M.; Mahon, C.; Haughton, C.; O'Neill, L.; O'Carroll, E. A consideration of the social impact of cybercrime: Examples from hacking, piracy, and child abuse material online. *Contemp. Soc. Sci.* **2015**, *11*, 373–391, doi:10.1080/21582041.2015.1117648. [CrossRef]
243. Ponemon Institute. 2017 Cost of Data Breach Study: Global Overview (Research Report). Ponemon Institute. 2017. Available online: https://www.ibm.com/downloads/cas/ZYKLN2E3 (accessed on 1 January 2020).
244. EC Horizon 2020, Secure Societies—Protecting Freedom and Security of Europe and Its Citizens. Available online: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/secure-societies-%E2%80%93-protecting-freedom-and-security-europe-and-its-citizens (accessed on 1 January 2020).
245. Bayerl, P.S.; Akhgar, B.; Brewster, B.; Domdouzis, K.; Gibson, H. Social media and its role for LEAs. In *Cyber Crime and Cyber Terrorism Investigator's Handbook*; Akhgar, B., Staniforth, A., Bosco, F., Eds.; Elsevier: Amsterdam, The Netherlands, 2014; pp. 197–220, doi:10.1016/B978-0-12-800743-3.00016-5. [CrossRef]
246. Donalds, C.; Osei-Bryson, K.-M. Toward a cybercrime classification ontology: A knowledge-based approach. *Comput. Hum. Behav.* **2019**, *92*, 403–418, doi:10.1016/j.chb.2018.11.039. [CrossRef]
247. Hicks, C. An ontological approach to misinformation: Quickly finding relevant information. In Proceedings of the 50th Hawaii International Conference on System Sciences, (HICSS 2017), Waikoloa Village, HI, USA, 4–7 January 2017; pp. 1–8.
248. Yu, F.; Liu, Q.; Wu, S.; Wang, L.; Tan, T. Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. *Comput. Secur.* **2019**, *83*, 106–121, doi:10.1016/j.cose.2019.02.003. [CrossRef]
249. Zhang, C.; Gupta, A.; Kauten, C.; Deokar, A.V.; Qin, X. Detecting fake news for reducing misinformation risks using analytics approaches. *Eur. J. Oper. Res.* **2019**, *279*, 1036–1052, doi:10.1016/j.ejor.2019.06.022. [CrossRef]
250. Shelke, S.; Attar, V. Source detection of rumor in social network—A review. *Online Soc. Netw. Media* **2019**, *9*, 30–42, doi:10.1016/j.osnem.2018.12.001. [CrossRef]
251. Bondielli, A.; Marcelloni, F. A Survey on fake news and rumour detection techniques. *Inf. Sci.* **2019**, *497*, 38–55, doi:10.1016/j.ins.2019.05.035. [CrossRef]

252. European Data Protection Supervisor. Meeting the Challenges of Big Data: A Call for Transparency, User Control, Data Protection by Design and Accountability, Opinion 7/2015. 2015. Available online: https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf (accessed on 1 January 2020).
253. Truyens, M.; van Eecke, P. Legal aspects of text mining. *Comput. Law Secur. Rev.* **2014**, *30*, 153–170, doi:10.1016/j.clsr.2014.01.009. [CrossRef]
254. Fatima, R.; Yasin, A.; Liu, L.; Wang, J.; Afzal, W.; Yasin, A. Sharing information online rationally: An observation of user privacy concerns and awareness using serious game. *J. Inf. Secur. Appl.* **2019**, *48*, 102351, doi:10.1016/j.jisa.2019.06.007. [CrossRef]
255. Chilton, P.A. *Analysing Political Discourse: Theory and Practice*; Routledge: London, UK, 2004.
256. Ludwig, S.A. MapReduce-based fuzzy c-means clustering algorithm: Implementation and scalability. *Int. J. Mach. Learn. Cybern.* **2015**, *6*, 923–934, doi:10.1007/s13042-015-0367-0. [CrossRef]
257. Kontopoulos, I.; Giannakopoulos, G.; Varlamis, I. Distributing n-gram graphs for classification. *Eur. Conf. Adv. Databases Inf. Syst.* **2017**, 3–11, doi:10.3389/fams.2018.00041. [CrossRef]
258. Paul, M.J.; Sarker, A.; Brownstein, J.S.; Nikfarjam, A.; Scotch, M.; Smith, K.L.; Gonzalez, G. Social media mining for public health monitoring and surveillance. In *Pacific Symposium on Biocomputing 2016, (PSB 2016)*; World Scientific Publishing Co.: Singapore, 2016; pp. 468–479, doi:10.1142/9789814749411_0043. [CrossRef]
259. Jordan, S.E.; Hovet, S.E.; Fung, I.C.-H.; Liang, H.; Fu, K.-W.; Tse, Z.T.H. Using Twitter for public health surveillance from monitoring and prediction to public response. *Data* **2018**, *4*, 6, doi:10.3390/data4010006. [CrossRef]
260. Lucini, F.R.; Fogliatto, F.S.; da Silveira, G.J.C.; Neyeloff, J.L.; Anzanello, M.J.; Kuchenbecker, R.S.; Schaan, B.D. Text mining approach to predict hospital admissions using early medical records from the emergency department. *Int. J. Med Inform.* **2017**, *100*, 1–8, doi:10.1016/j.ijmedinf.2017.01.001. [CrossRef] [PubMed]
261. Metsker, O.; Bolgova, E.; Yakovlev, A.; Funkner, A.; Kovalchuk, S. Pattern-based mining in electronic health records for complex clinical process analysis. *Procedia Comput. Sci.* **2017**, *119*, 197–206, doi:10.1016/j.procs.2017.11.177. [CrossRef]
262. Leong, C.K.; Lee, Y.H.; Mak, W.K. Mining sentiments in SMS texts for teaching evaluation. *Expert Syst. Appl.* **2012**, *39*, 2584–2589, doi:10.1016/j.eswa.2011.08.113. [CrossRef]
263. He, W. Examining students' online interaction in a live video streaming environment using data mining and text mining. *Comput. Hum. Behav.* **2013**, *29*, 90–102, doi:10.1016/j.chb.2012.07.020 [CrossRef]
264. Rodrigues, M.W.; Isotani, S.; Zárate, L.E. Educational data mining: A review of evaluation process in the e-learning. *Telemat. Inform.* **2018**, *35*, 1701–1717, doi:10.1016/j.tele.2018.04.015. [CrossRef]
265. Ferreira-Mello, R.; André, M.; Pinheiro, A.; Costa, E.; Romero, C. Text mining in education. *WIREs Data Min. Knowl. Discov.* **2019**, e1332, doi:10.1002/widm.1332. [CrossRef]
266. Zaeem, R.; Manoharan, M.; Yang, Y.; Barber, K.S. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Comput. Secur.* **2017**, *65*, 50–63, doi:10.1016/j.cose.2016.11.002. [CrossRef]
267. Das, P.; Das, A.K. Graph-based clustering of extracted paraphrases for labelling crime reports. *Knowl. Based Syst.* **2019**, *179*, 55–76, doi:10.1016/j.knosys.2019.05.004. [CrossRef]
268. Amrit, C.; Paauw, T.; Aly, R.; Lavric, M. Identifying child abuse through text mining and machine learning. *Expert Syst. Appl.* **2017**, *88*, 402–418, doi:10.1016/j.eswa.2017.06.035. [CrossRef]
269. Esser, D.E.; Williams, B.J. Tracing poverty and inequality in international development discourses: An algorithmic and visual analysis of agencies' annual reports and occasional white papers, 1978–2010. *J. Soc. Policy* **2014**, *43*, 173–200, doi:10.1017/S0047279413000342. [CrossRef]



big data and cognitive computing

MDPI

Concept Paper

Seven Properties of Self-Organization in the Human Brain

Birgitta Dresp-Langley

ICube Lab UMR 7357 Centre National de la Recherche Scientifique, University of Strasbourg, 67085 Strasbourg, France; birgitta.dresp@unistra.fr

Received: 24 March 2020; Accepted: 8 May 2020; Published: 10 May 2020

Abstract: The principle of self-organization has acquired a fundamental significance in the newly emerging field of computational philosophy. Self-organizing systems have been described in various domains in science and philosophy including physics, neuroscience, biology and medicine, ecology, and sociology. While system architecture and their general purpose may depend on domain-specific concepts and definitions, there are (at least) seven key properties of self-organization clearly identified in brain systems: (1) modular connectivity, (2) unsupervised learning, (3) adaptive ability, (4) functional resiliency, (5) functional plasticity, (6) from-local-to-global functional organization, and (7) dynamic system growth. These are defined here in the light of insight from neurobiology, cognitive neuroscience and Adaptive Resonance Theory (ART), and physics to show that self-organization achieves stability and functional plasticity while minimizing structural system complexity. A specific example informed by empirical research is discussed to illustrate how modularity, adaptive learning, and dynamic network growth enable stable yet plastic somatosensory representation for human grip force control. Implications for the design of "strong" artificial intelligence in robotics are brought forward.

Keywords: self-organization; computational philosophy; brain; synaptic learning; adaptation; functional plasticity; activity-dependent resonance states; circular causality; somatosensory representation; prehensile synergies; robotics

1. Introduction

The principle of self-organization [1] governs both structure and function, which co-evolve in self-organizing systems. Self-organizing systems [2] differ from any computational system where the architecture and all its functional aspects are created and controlled by their designer. In line with previous attempts at a comprehensive definition of the concept [1], the author proposes that self-organization may be defined in terms of a general principle of functional organization that ensures a system's auto-regulation, stability, adaptation to new constraints, and functional autonomy. A self-organizing system [2], beyond the fact that it regulates and adapts its own behavior [2,3], is capable of creating its own functional organization [2–4]. The concept of self-organization acquires a critically important place in the newly emerging field of computational philosophy, where it inspires and lends conceptual support to new approaches to complex problems, in particular in the field of Artificial Intelligence (AI) [5].

Mathematical developments evoking self-organization as a general functional principle of learning and adaptation in biological systems hark back to Hebb's work on synaptic plasticity [6] and to work by Minsky and colleagues [7] at the dawn of research on artificial intelligence in the context of Rosenblatt's PERCEPTRON model [8]. Self-organization is the foundation of what is sometimes referred to as "strong AI" [5]. Systems with self-organizing properties [2] have been developed in physics [9], ecology and sociology [10,11], biology and medicine [12], and in neuroscience [3,13] and perceptual neuroscience [3,14] in continuity with the earlier PERCEPTRON approaches. Structure and functional

organization of self-organizing systems vary depending on the field. Their properties relate to function more than to components. The fields of neuroscience and artificial intelligence in particular share a history of interaction in the theoretical development of both the concept of self-organization and self-organizing systems, and many of the current advances in AI were inspired by the study of neural processes in humans and other living species [3,5,13].

Neuroscience provides a source of inspiration for new algorithms and architectures, independent of and complementary to mathematical methods. Such inspiration is well-reflected by many of the concepts and ideas that have largely dominated traditional approaches to AI. Neuroscience may also convey external validity to AI. If an algorithm turns out to be a good model for a functionally identified process or mechanism in the brain, then such biological plausibility lends strong support to the fitness of the algorithm for the design of an intelligent system. Neuroscience may thus help conceive new algorithms, architectures, functions, and codes of representation for the design of biologically plausible AI by using a way of thinking about similarities and analogies between natural and artificial intelligence [15]. Such two-way conceptual processes acquire a particular importance in the newly emerging field of computational philosophy, which regroups a wide range of approaches relating to all fields of science.

Computational philosophy [16] is aimed at applying computational techniques, models, and concepts to advance philosophical and scientific discovery, exploration, and argument. Computational philosophy is neither the philosophy of computation, an area that asks about the nature of computation itself, nor the philosophy of artificial intelligence. Computational philosophy represents a self-sufficient area with a widespread application across the full range of scientific and philosophical domains. Topics explored in computational philosophy may draw from standard computer programming, software engineering, artificial intelligence, neural networks, systems science, complex adaptive systems, and computer modeling. As a relatively young and still growing domain, its field of application is broad and unrestricted within the traditional discipline of general philosophy. In the times of Newton, there was no epistemological boundary between philosophy and science. Across the history of science, there has never been a clear division between either computational and non-computational philosophy, or computational philosophy and other computational disciplines [16].

The place of computational philosophy in science entirely depends on the viewpoint adopted and goal pursued by the investigator [17,18]. If the goal pursued is to enrich computational philosophy based on an understanding of brain processes, then the functional characteristics of brain mechanisms may fuel the development of computational philosophy. If the goal is to enrich brain science based on computational philosophy, then empirical brain research will be fueled by computational philosophy for building brain models reflective of mechanisms identified in the human brain [17,18]. This article is a conceptual essay written from the viewpoint of computational philosophy. It highlights seven general functional key properties related to the principle of self-organization, which are then discussed under the light of a specific example from sensory neuroscience, backed by empirical data. How modularity, adaptive learning, and dynamic network growth enable stable somato-sensory representation for human grip force control in a biological neural network (*from hand to brain and back*) with previously identified functional plasticity is illustrated.

Since structure and functional organization co-evolve in self-organizing systems [1,2], one cannot account for such systems without providing an account for the functional properties most closely linked to its self-organizing capacity. The latter is generally described in terms of spatiotemporal synergies [1,9]. In the brain, neurons respond at time scales of milliseconds, while perception, which is experience and memory dependent, takes longer to form. Such timescale separation between long-time scale parameters and short-time scale functioning [19,20] is akin to that described for physical synergetic systems [9] and reflects the circular causality that is characteristic of self-organization in general [1]. The following sections start with an overview of seven key properties of systems that "self-organize". This is followed by a discussion of examples of such properties in the human somato-sensory system [21,22] involved in the control of prehensile synergies for grip-force adaptation. The example

provides a biologically plausible conceptual support for the design of autonomous self-organization (AI) in soft robotics and illustrates why the seven key properties brought forward here in this concept paper are conducive to advancing the development of "strong AI" [5], as pin-pointed in the conclusions.

2. Seven Key Properties of Self-Organization

Seven properties linked to the principle of self-organization have been described on the basis of functional investigation of the human brain: (1) modular connectivity, (2) unsupervised learning, (3) adaptive ability, (4) functional resiliency, (5) functional plasticity (6) from-local-to-global functional organization, and (7) dynamic system growth.

2.1. Modular Functional Architecture and Connectivity

Modularity refers to a computational and/or structural design principle for systems that can be decomposed into interacting subsystems (nodes, modules) that can be understood independently. Modular systems design is aimed at reducing complexity [23] by a fundamental design principle identified in biological neuronal systems at the scale of cells (units, neurons), local circuits (nodes), and interconnected brain areas (subsystems) [24]. The human brain's neuronal network architecture is not based on a genetically preformatted design, although some of it may be prewired, but is progressively shaped during ontogenetic development by physiological and chemical changes that obey computational rules of activity-dependent self-organization [25]. At the medium level of local circuits, the brain (cortex) is organized in local clusters of tightly interconnected neurons that share common input. The neuronal targets that constitute a basic computational module share similar functional properties. Activity-dependent self-organization influences the system's modularity on the one hand, and modular connectivity promotes spontaneous firing activity on the other [25]. Thus, the modular connectivity of a self-organizing system and its capacity for self-organization are interdependent, and they co-evolve in a mutually reinforcing process to ensure the simultaneous development of both structural and functional capacity. This entails that the more such a system learns, the more active connections it will develop on the one hand, and the more it will be able to learn, on the other. For its structural and functional development, a self-organizing system exclusively uses unsupervised learning.

2.2. Unsupervised Learning

Unsupervised learning [6,26–28], one may also call it self-reinforced learning, is essential to the principle of self-organization, as illustrated by functional dynamics of the neural network systems described in Adaptive Resonance Theory (ART) and Self-Organizing Maps (SOM). Unsupervised learning is essential to overcome the stability–plasticity dilemma in neural networks, and both ART [27] and SOM [28,29] are based on unsupervised approaches that are fundamental in machine learning in general and in Artificial Intelligence (AI) in particular. The Hebbian synapse and synaptic learning rules [6] are the fundamental conceptual basis of unsupervised learning in biological [30] and artificial neural networks [31]. A synapse refers to connection between two neurons in a biological or artificial neural network, where the neuron transmitting information via a synapse or synaptic connection is referred to as the pre-synaptic neuron, and the neuron receiving the information at the other end of a synaptic connection as the post-synaptic neuron (Figure 1). The information propagation efficiency of biological and artificial synapses is strictly self-reinforcing as the more a synapse is stimulated, the more effectively information flows through the connection, which ultimately results in what Hebb [6] and subsequently others [3,31] have called Long-Term Potentiation (LTP). Synaptic connections that are not repeatedly stimulated and as a consequence not self-reinforced will lose their information propagation efficiency, which ultimately results in Long-Term Depression (LTD).

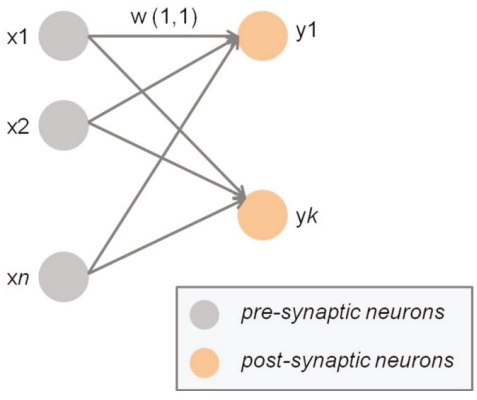


Figure 1. Schematic illustration of Hebbian synapses within a small neural network. Self-reinforcing synaptic learning is by definition unsupervised and involves the progressive increment of the synaptic weights (w) of efficiently stimulated connections, which are thereby long-term potentiated, while non-reinforced synapses will lose their efficiency and ultimately become long-term depressed.

The information propagation in unsupervised synaptic learning in neural networks may be event-driven [32], clock-driven [33], or a combination of both [31]. The Long-Term Potentiation (LTP) of efficient synaptic connections on the one hand, and the Long-Term Depression (LTD) of inefficient connections on the other promote the emergence of an increasingly effective functional organization in the neural network akin to that found in biological organisms, chemical structures, and ecosystems. In computational thinking and philosophy, self-reinforcing Hebbian learning may, indeed, be seen as a ground condition for adaptive system function, a highly dynamic, unsupervised learning process where local connections change state towards potentiation or depression, depending on how efficiently they propagate information across the system. Synaptic long-term potentiation and long-term depression are not definitive. A depressed state may reverse to a potentiated one, and vice versa, as a function of a persistent change in the system's external environment (stimuli), or of a specific chemical treatment or drug in the case of biological neural networks. LTP and LTD have acquired a potentially important role in contemporary neuroscience [34] and may be exploited to treat disorder and disease in the human brain, knowing that a variety of neurological conditions arise from either lost or excessive synaptic drive due to sensory deprivation during childhood, brain lesions, or disease [35]. Manipulation of relative synaptic efficiency using various technologies may provide a means of normalizing synaptic strength and thereby ameliorating plasticity-related disorders of the brain [34,35]. Thus, it may indeed be argued that the reversibility of synaptic efficiency as a function of changes during self-reinforcing learning drives the self-organizing system's adaptive ability.

2.3. Adaptive Ability

A self-organizing system will adapt its functional organization to significant changes in the environment by the ability to learn new "tricks" to cope with new problems of increasing complexity. Biological neural networks in the human central nervous system have the ability to adapt their functional fine-tuning to sudden changes in the environment [36], making the brain remarkably efficient at coping with an often unpredictable, ever-changing external world. In any self-organizing neural network, such adaptation relies heavily on their modular connectivity on the one hand, and on synaptic (Hebbian) plasticity, on the other [37–39]. Research on so-called neuromodulation [40–42], a biological mechanism that dynamically controls intrinsic properties of neurons and their response to external stimuli in a context-dependent manner, has produced simulations of adaptive system behavior. Adaptive Resonance Theory (ART) uses self-organization to explain how normal and abnormal brains learn to categorize and recognize objects and events in a changing world, and how

normal as well as abnormal learned categories may be consolidated in memory for a long time [42]. Thus, brain pathology may be conceived in terms of a self-organizing system's adaptive ability gone wrong. This possibility was already to some extent taken into consideration by Darwin [43], who noted that natural selection can act together with other processes, including random changes in the frequencies of phenotypic differences that are not under strong selection, and changes in the environment, which may reflect evolutionary changes in the organisms themselves. At the cellular level, adaptation refers to changes in a cell's or neuron's behavior in response to adverse or varying environmental changes. Adaptation may be normal, or pathological (abnormal), depending on extent and type of environmental pressure on the cell, a system of cells, or a whole brain network [44–47]. Under extreme external conditions, some forms of pathological adaptation may be an effective means towards the general goal of survival. In human behavior, the Stockholm Syndrome is one such example, where hostages start taking sides with their aggressors to functionally adapt to the terrible fact that they are at their complete mercy. Adaptive system ability as a concept, in the brain sciences and in computational philosophy, helps conceive intelligent systems with a capacity to generate order from, or preserve order within, external chaos. Seemingly random perturbations will help a self-organizing system develop and perform even better, rather than prevent its evolution. Perturbations will promote the emergence of an increasingly effective functional organization, as found in biological organisms, chemical structures, or ecosystems. Self-organizing adaptation is to be seen as a highly dynamic process [20,46,47], where components are constantly changing state as a function of state changes in other components. Such complex mutual dependency in self-organization was not known in the times of Darwin, and is therefore not included in traditional definitions of the concept "adaptation" [43–45]. The adaptive systemic changes we are talking about here in this concept paper are determined by self-reinforcement of connections that profit the system's functioning and ensure its dynamic functional growth on the one hand, and by local suppression of connections that are either redundant, or disserve the system, on the other. Both processes, reinforcement and inhibition, are critical to sustain a self-organizing system's ability to cope with unexpected external changes or pressure, and thereby also ensure its functional resiliency.

2.4. Functional Resiliency

A self-organizing system's adaptive ability implies functional resiliency. After lesion or damage, a human brain will continue to function, often astonishingly well and without any detectable change in efficiency. The human brain can endure numerous micro-strokes with seemingly no detrimental impact and is resilient against both targeted and random damage or lesions [48,49]. Self-organizing systems, like the human brain, are intrinsically robust and can withstand a variety of perturbations, even partial destruction. The strength of functional interaction between any two system nodes is not solely determined by the presence or absence of a direct connection, but mostly by the number of indirect (long-range) connections [50]. These long-range connections, which will be discussed in greater detail in 5) here below, are indispensable to ensure self-repair or self-correction of partial systemic damage and make the system capable of returning to its initial functional state after local damage. By virtue of their modular connectivity discussed here above in 1), and self-reinforcing learning capacity discussed here above in 2), system components or subsystems (synapses or networks) that have initially learnt to fulfill a specific function can spontaneously adapt to perform a different new function that was previously ensured by the damaged component(s). This self-organizing ability is referred to as functional plasticity.

2.5. Functional Plasticity

The functional resiliency of a self-organizing system implies functional plasticity [21,22,51], which is a necessary ground condition for system resiliency, but also achieves a purpose well beyond this. Functional plasticity ensures system functioning under adverse conditions and/or after partial system damage. Posttraumatic stress disorder (PTSD), for example, is associated with

plastic functional changes in the human medial prefrontal cortex, hippocampus, and amygdala that correlate with a smaller hippocampal volume, and both reversed to normal after treatment [52]. Like in the human brain, where a functional subsystem may take over the functions of another after brain damage [21,22,51], a functional subsystem may appear spontaneously and maintain its function autonomously by self-organization in a computer generated system. The control needed to achieve this has to be distributed across system levels, components or cells, and/or sub-systems. If system control were centralized in a subsystem or module, then the system as a whole would lose its organization whenever the sub-system is damaged or destroyed. Use-dependent long-term changes of neuronal response properties must be gated to prevent irrelevant activity from inducing inappropriate modifications. Local network dynamics contribute to such gating, as synaptic modifications [53] depend on temporal contiguity between pre-synaptic and post-synaptic activity, there are observable stimulation-dependent modifications, as shown on the example of orientation selectivity in adult cat visual cortex [35]. The stability-plasticity dilemma, a constraint for intelligent systems, is potentially resolved in self-organizing systems, such as those in ART [3,14,47]. Plasticity is necessary for the integration of new knowledge by self-reinforced learning, but too much of it compromises systemic stability and may cause catastrophic forgetting of previous knowledge [54]. It is assumed that too much plasticity will result in previously learnt data being constantly forgotten, whereas too much stability will hinder self-reinforced learning at the synaptic level, yet, the exact functional relationship between changes in synaptic efficacy and structural plasticity is not entirely understood. It has been proposed that a continuum exists between the two, such that changes in synaptic efficacy precede and instruct structural changes, however, in other cases, structural changes may occur without any stimulation producing an initial change in local synaptic efficiency [55], which points towards the critical functional role of long-range connections [56] within the from-local-to-global functional organization of self-organizing systems.

2.6. From-Local-to-Global Functional Organization

In a self-organizing neural network, changes in the system during self-reinforced synaptic learning are initially local, as components or neurons initially only interact with their nearest "neighbors". Though local connections are initially independent of connections farther away, self-organization generates "global order" on the basis of many spontaneous, initially local, interactions [56], where the most efficient synaptic connections self-reinforce, are long-term potentiated as described here above in 2) and, ultimately, acquire propagation capacity beyond local connections. This leads to the formation of functionally specified long-range connections, or circuits which, by virtue of self-organization, self-reinforce on the basis of the same Hebbian principles that apply to single synapses; however, the rules by which long-range circuits of connections learn can no longer be accounted for in terms of a Hebbian linear model. The human brain is, again, the choice example of a complex biological structure where local, modular processing potentiates global integrative processing. Current functional brain anatomy suggests areas that form domain-specific hierarchical connections [57,58] on the one hand, and multimodal association areas receiving projections from more widely distributed functional subsystems [59]. Dominance of one connectivity profile over the other can be identified for many areas [56], revealing the self-organizing principles of long-range cortical-cortical functional connectivity. Early visual cortical areas such as V1 and V2 already show a functional organization beyond strictly local hierarchical connections [58,60]. The prefrontal, temporal, and limbic areas display "functional hubs" [56], projecting long-range connections across larger distances to form the "neural epicenters" [56] of scale-free, distributed brain networks [61,62]. The "beyond the classic receptive field" functional organization of the visual brain has been progressively unraveled in behavioral and functional neuroscience over the last 30 years [60]. The discovery of input effects from beyond the "classic receptive field", as previously identified and functionally defined in much earlier, Nobel prize awarded work [63–66], has shown that neuronal activity recorded from cortical areas V1 and V2 in response to visual stimuli is modulated by stimuli presented outside the corresponding

receptive fields on the retina [67–69]. This is direct evidence for contextual modulation of neural activity and indirectly reflects functional properties of long-range neural connections at early processing levels in the visual brain [68,69]. In higher visual areas such as the temporal lobe, visual receptive fields increase in size and lose retinotopic organization, encoding increasingly complex features [60]. This from-simple-to-complex, self-organizing functional hierarchy forms the core of the LAMINART model family [70,71] of Adaptive Resonance Theory [3,14,47]. In the LAMINART neural networks, self-organizing long-range cooperation and short-range competition, whereby locally stimulated bipolar neurons complete boundaries across gaps in oriented line or edge, contrast stimuli by receiving strong excitatory inputs from both sides, or just one side of their receptive fields. The more strongly activated bipolar cells inhibit surrounding bipolar cells within a spatially short-range competitive network. The short-range network communicates with long-range resonant feedback networks connecting the interblob and blob cortical streams within V1, V2, and V4 of the visual cortex. The resonant feedback networks enable boundaries and surfaces in images to emerge as consistent representations on the basis of computationally complementary rules. This self-organizing property of resonant feedback networks in ART is called complementary consistency [72]; the computational mechanisms that ensure complementary consistency contribute to three-dimensional perceptual organization [72–75]. The long-range resonant properties of the neural network architectures exploited by ART enable the self-organizing system to grow dynamically.

2.7. Dynamic Functional Growth

A self-organizing system is dynamic and its components (cells, neurons, circuits) are constantly changing states relative to each other. As explained here above, under 1), structure and function of the system are mutually dependent, which entails that the changes that occur while such a system is developing further, i.e., growing, are not arbitrary but activity-dependent [76–79]. While the system grows by changing states, there will be relative states that will be particularly beneficial to the system's effectiveness and, as a consequence, these states self-reinforce along similar principles as those described here above in 2). When consistently reinforced, newly emerging system states will, ultimately, become stable states, but with functional plasticity as explained here above in 4) to resolve the stability-plasticity dilemma [3,54]. Less beneficial or useless new relative states will not self-reinforce and, as a consequence, be inhibited and ultimately functionally depressed. Each connection within a self-organizing system has its own, individual characteristics, like a species within an ecosystem. A particular example of non-linear dynamic functional growth in fixed-size neural networks (Figure 2) would be activity-dependent formation of dedicated resonant circuitry [3,79].

The amount of different individual characteristics within the functional structure directly determines the system's functional complexity. Individual components, or cells, fit in a functionally specified "niche" within the system. The propagation of fits is self-reinforcing by nature, and the larger the niche, the quicker the propagation of additional functions in increasingly larger sub-circuits, exerting increasingly stronger attraction on functionally still independent cells. Such propagation of fits drives a positive feedback process enabling explosive system growth, and the production of new functional circuitry in existing network structures. A system's growth generally stops when the system resources are exhausted. A self-organizing system, however, never stops growing dynamically. As illustrated here above on the example of resonant sub-circuitry formation in structurally fixed/limited neural networks (Figure 2), self-organized functional growth does not require adding structural component (cells, neurons, subsystems). The system can grow and develop new, dynamic functionalities without the need for adding further structural complexity. As the external environment of the system changes, functional components or cells directly interacting with the environment will adapt their state(s) in a self-reinforcing process to maintain their fitness within the system. This adaptive fit will propagate further inwards, until the whole functional structure is fully adapted to the new situation. Thus, a dynamically growing self-organizing system constantly re-organizes by mutually balancing internal and external pressures for change while trying to maintain

its general functional organization and to counteract any loss thereof. Functional self-preservation is, indeed, a self-organizing system's main purpose, and each component or cell is adaptively tuned to perform towards this goal. A self-organized system is stable, largely scale-invariant, and robust against adverse conditions. At the same time, it is highly dynamic. In physics, systems achieve so-called "criticality" by the fine-tuning of control parameters to a specific value for which system variations become scale-invariant. In biological systems, criticality occurs without the need for such fine-tuning. The human brain is an example of such a system. This self-tuning to criticality, accounted for in physics by graph theory, is called self-organized criticality [80].

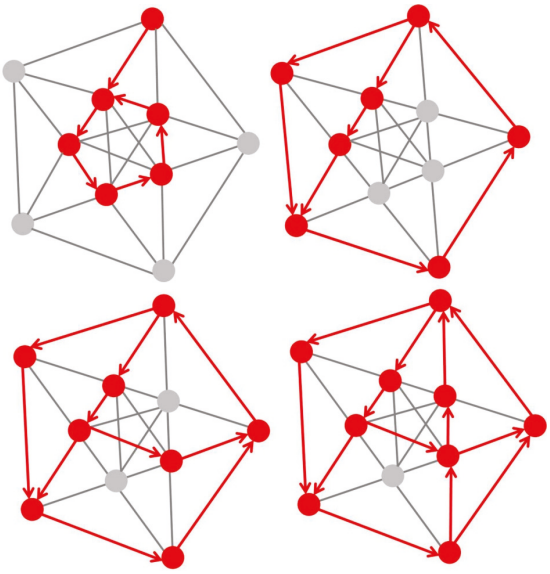


Figure 2. Schematic illustration of self-organizing formation of dedicated resonant circuitry in a fixed-size neural network of 10 neurons only. An exponential number of functional states are possible therein, allowing even a small-size network to develop new, dynamic functionalities without the need for adding further structural complexity. Resonant neurons (highlighted in red color) are primed throughout their functional development to preferentially process input which carries statistically "strong" signals, as explained previously in [79]. When activated, resonant neurons send signals along all delay paths originating from them, and all those receiving a signal coinciding with the next input signal remain activated. The formation of resonant circuitry is activity-dependent; long-term potentiated connections between resonant neurons may become progressively depressed, as in the model of self-reinforcing (Hebbian) synaptic learning, when their function is no longer activated.

3. Seven Properties of Self-Organization in a Somatosensory Neural Network

The brain structures that subserve cognitive functions require sensory experience for the formation of neuronal connections by self-organization during learning [81]. Neuronal activity and the development of functionally specific neural networks in the continuously learning brain are thus modulated by sensory signals. The somatosensory cortical network [82], or S1 map, in the primate brain is an example of such self-organization. S1 refers to a neocortical area that responds primarily to tactile stimulations on the skin or hair. Somatosensory neurons have the smallest receptive fields and receive the shortest-latency input from the receptor periphery. The S1 cortical area is conceptualized in current state of the art [82,83] as containing a single map of the receptor periphery. The somatosensory cortical network has a modular functional architecture and connectivity (property 1), with highly specific connectivity patterns [82–85], binding functionally distinct neuronal

subpopulations from other cortical areas into motor circuit modules at several hierarchical levels [84]. The functional modules display a hierarchy of interleaved circuits connecting via inter-neurons in the spinal cord, in visual sensory areas, and in motor cortex with feed-back loops, and bilateral communication with supraspinal centers [84,85]. The from-local-to-global functional organization (property 6) of motor circuits relates to precise connectivity patterns, and these patterns frequently correlate with specific behavioral functions of motor output. Current state of the art suggest that developmental specification, where neuronal subpopulations are specified in a process of precisely timed neurogenesis [85], determines the self-organizing nature of this connectivity for motor control, in particular limb movement control [84,85]. The functional plasticity (property 5) of the somatosensory cortical network is revealed by neuroscience research investigating the somatosensory cortical map has shown that brain representations change adaptively following digit amputation in adult monkeys [21]. In the human primate [86], somatosensory representations of the fingers left intact after amputation of others on the same hand become expanded in less than 10 days after amputation, when compared with representations in the intact hand of the same patient, or to representations in either hand of controls. Such network expansion reflects the functional resiliency (property 4) of the self-organized somatosensory system.

The human hand has evolved [87] as a function of active constraints [88–100] and is in harmony with other sensory systems such as the visual and auditory brain [97,99,101]. Grip force profiles are a direct reflection of complex low-level, cognitive, and behavioral synergies this evolution has produced [87–101]. The state of the art in experimental studies on grip force control for lifting and manipulating objects [88,89,91,93,94] provides insight into the contributions of each finger to overall grip strength and fine grip force control. The middle finger, for example, has evolved to become the most important contributor to gross total grip force and, therefore, is most important for getting a good grip of heavy objects to lift or carry, while the ring finger and the small (pinky) finger have evolved for the fine control of subtle grip force modulations, which is important in precision tasks [102–106]. Human grip force is governed by self-organizing prehensile synergies [91,92] that involve from-local-to-global functional interactions (property 6) between sensory (low-level) and central (high-level) representations in the somatosensory brain. Grip force can be stronger in the dominant hand compared with the non-dominant hand and may reverse spontaneously depending on the necessity for adaptive ability (property 3) as a function of specific environmental constraints [95,96,100]. In recent studies, the grip force profiles from thousands of force sensor measurements collected from specific locations on anatomically relevant finger parts on the dominant and non-dominant hands revealed spontaneous adaptive grip force changes in response to sensory stimuli [105], and long-term functional plasticity (property 5) as a function of task expertise [103,104].

Somatosensory cortical neural networks of the S1 map [81] develop their functional connectivity [83–86] through a self-organizing process of activity-dependent, dynamic functional growth (property 7). This process is fueled by unsupervised learning (property 2) and, more specifically, synaptic (cf. Hebbian) learning, which drives spontaneous functional adaptation as well as long-term functional re-organization and plasticity [22,23,54,56,83]. An example of this process, fueled by recent empirical data from thousands of sensor data collected from anatomically relevant locations in the dominant and non-dominant hands of human adults, is illustrated here in Figures 3–5.

The grip force profiles of a novice (beginner) and a skilled expert in the manipulation of the robotic device directly reflect such differences in somatosensory cortical representation before and after learning. Individual grip force profiles, corresponding to thousands of individual force sensor data, were recorded in real time from different sensor locations, including R1 and R2 (Figure 3) on the middle phalanxes of the small (R1) and the middle fingers (R2) in the hands of a total beginner and the expert across several robotic task sessions. The grip force profiles are shown in Figure 4 here below. They display two radically different functional states with respect to gross (middle finger) and fine (small finger) grip force control, translating motor expertise and functionally reorganized somatosensory network states driven by self-organization.

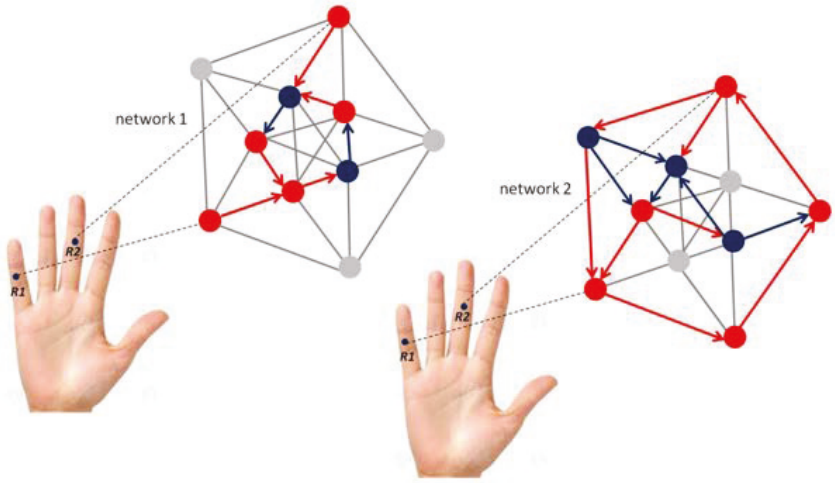


Figure 3. Schematic illustration of self-organizing functional reorganization in a fixed-size neural network in the somatosensory brain before (left) and after (right) unsupervised learning of a visually guided manual robotic precision task [103,104]. Mechanoreceptors on the middle phalanxes of the small (R1) and the middle fingers (R2) are indicated. Two different network representations correspond to brain-behaviour states before (left) and after acquisition of grip force expertise (right) for performing the robotic precision task. The different levels of connectivity used here are arbitrarily chosen, and for illustration only; single nodes in the networks displayed graphically here may correspond to single neurons, or to a subpopulation of neurons with the same functional role. Red nodes may represent motor cortex (M) neurons, blue nodes may represent connecting visual neurons (V). Only one-way propagation is shown here to keep the graphics simple, knowing that the somatosensory brain has multiple two-way propagation pathways with functional feed-back loops [82–85].

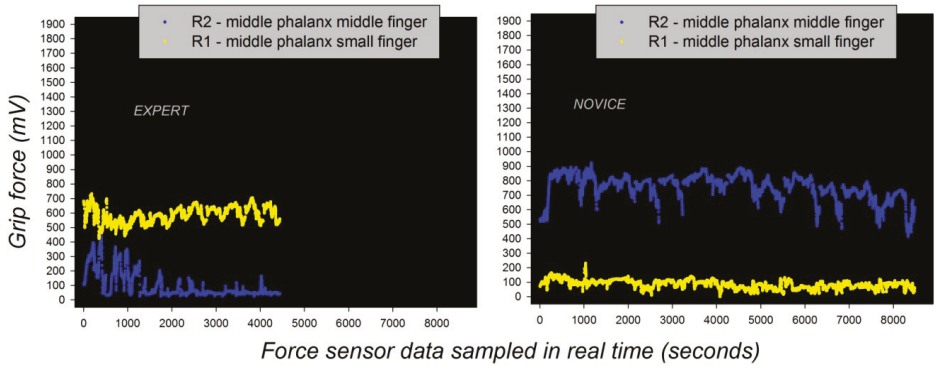


Figure 4. Individual grip force profiles of an expert (left) and a beginner (right) in a robotic task [103,104] reflecting two radically different functional states with respect to gross (middle finger) and fine (small finger) grip force control, translating motor expertise and functionally reorganized somatosensory network states (cf. Figure 3), governed by self-organization.

The self-organized functional reorganization due to plasticity shown here is stable, as reflected by stable grip force profiles in the expert across task sessions [103]. By comparison, the grip force profiles of the beginner do not display the same stability, as shown by the statistical analyses reported elsewhere [103,104]. It may be assumed that stable dynamic functional growth in the neural

network system generating the somatosensory representations is driven by such long-term plasticity, which reflects a process of long-term adaptation to specific task constraints. This long-term adaptation ensures system stability, but does not reflect a permanent system state. However, the somatosensory system also displays spontaneous functional plasticity and reorganization and rapid adaptation to new constraints [21,22,81–85].

The grip force profiles of one and the same individual adapt spontaneously to new sensory input from other modalities, as predicted by the general functional organization of the somatosensory brain networks [82–86]. Such spontaneous adaptive ability is reflected by dynamic changes that are short-term potentiated, rather than reflective of long-term plasticity. An example of spontaneous grip force adaptation to new visual input in one and the same individual is shown here below in Figure 5. The subject was blindfolded first, then made to see again, during a bimanual grip task where young male adults had to move two weighted handles up and down [105]. The individual grip force profiles corresponding to force sensor recordings from the middle phalanxes of the forefinger and the middle finger in the dominant hand are shown for comparison (Figure 5).

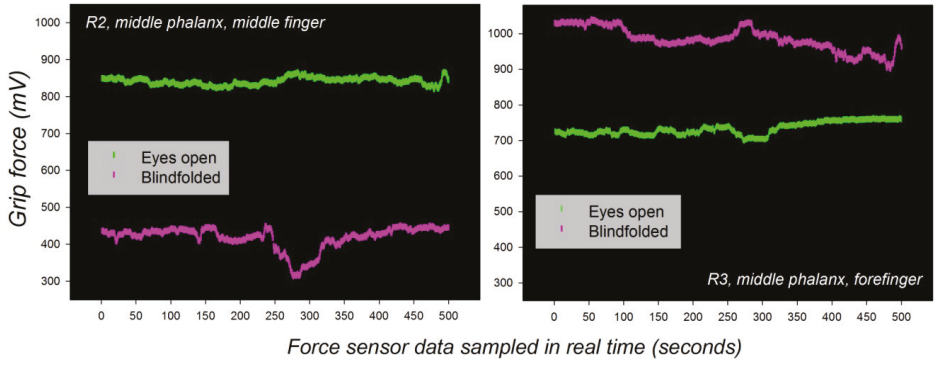


Figure 5. Spontaneous adaptive reorganization of an individual's grip force profile (dominant hand) during a bimanual grip task executed first blindfolded, and then with both eyes open. Unpublished data from a study described in [106] are shown. Gross grip force in the middle finger spontaneously increases with sudden visual input (left), while simultaneous grip force in the forefinger decreases (right).

The example here above illustrates seven key properties of self-organization in the somatosensory brain and points towards the implications of self-organized brain learning for the design of robust control schemes in large complex systems with unknown dynamics, which are difficult to model [107,108]. Beyond the functional stability and resilience of self-organizing systems, self-reinforced unsupervised learning based on differential plasticity, with feedback control through internal system dynamics, may enable robots [107] to learn to relate objects on the basis of specific sensorimotor representations (Figure 6), for example.

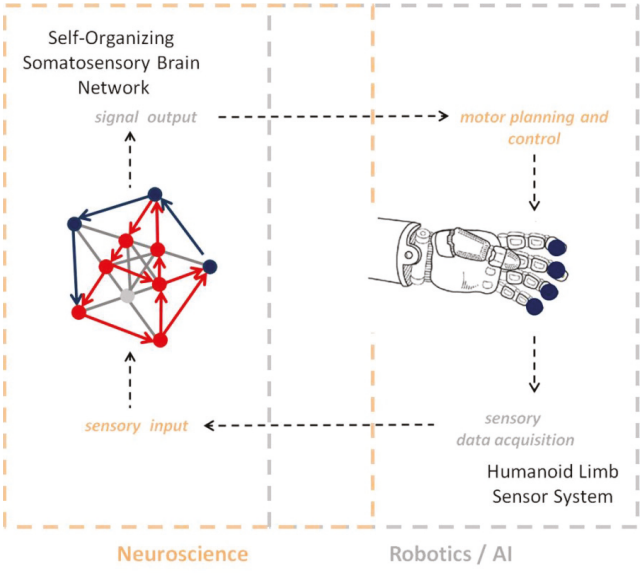


Figure 6. Conceptual diagram illustrating a "closed-loop" link between systems neuroscience and robotics/AI.

4. Conclusions

Self-organization is a major functional principle that allows us to understand how the neural networks of the human brain continuously generate new knowledge at all hierarchical levels, from sensory to cognitive representation. From the philosophical standpoint, the principle of self-organization establishes a clear functional link between the "mental" and the "physical" [109,110]. As a fundamental conceptual support for the design of intelligent systems, it provides conceptual as well as mathematical [1–3] tools to help achieve system stability and reliability, while minimizing system complexity. It enables specific fields such as robotics to conceive new, adaptive solutions based on self-reinforced systemic learning, where the activity of connections directly determines their performance and, beyond robotics, allows for the conceptual design of a whole variety of adaptive systems that are able to learn and grow independently without the need for adding non-necessary structural complexity. There is a right balance between structural and functional complexity, and this balance conveys functional system plasticity; adaptive learning allows the system to stabilize but, at the same time, remain functionally dynamic and able to learn new data. Activity-dependent functional systemic growth in minimalistic sized network structures is probably the strongest advantage of self-organization; it reduces structural complexity to a minimum and promotes dimensionality reduction [111,112], which is a fundamental quality in the design of "strong" Artificial Intelligence [5]. Bigger neural networks akin to those currently used for deep learning [113] do not necessarily learn better or perform better [114]. Self-organization is the key to designing networks that will learn increasingly larger amounts of data increasingly faster as they learn, consolidate what has been learnt, and generate output that is predictive [47,68] instead of being just accurate. In this respect, the principle of self-organization will help design Artificial Intelligence that is not only reliable, but also meets the principle of scientific parsimony, where complexity is minimized and functionality optimized. For example, a single session of robot controlled proprioceptive training induces connectivity changes in the somatosensory networks associated with residual motor and sensory function [115], translating into improved motor function in stroke patients. This example perfectly illustrates the closed-loop epistemological link (Figure 6) between systems neuroscience

and robotics/AI. Robotic sensory learning models inspired by self-organizing somatosensory network dynamics (in short: AI) are currently developed [115–117], within and well beyond the context of motor rehabilitation programs. Further studies on the effects of repeated training sessions on neural network learning and somatosensory functional plasticity will 1) enable the possible generalization of motor relearning and treatment effects in the clinical domain and 2) the development of reliable robot motor learning and control on the basis of "strong" neuro-inspired AI.

Author Contributions: The author is the lead for all aspects of this research. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The support of the CNRS is gratefully acknowledged.

Conflicts of Interest: The author declares no conflict of interest.

References

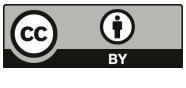
1. Haken, H. Self-Organization. *Scholarpedia* **2008**, *3*, 1401. [CrossRef]
2. Self-Organizing Systems. Science Direct Topics. Available online: https://www.sciencedirect.com/topics/physics-and-astronomy/self-organizing-systems (accessed on 29 April 2020).
3. Grossberg, S. Self-organizing neural networks for stable control of autonomous behavior in a changing world. In *Mathematical Approaches to Neural Networks*; Taylor, J.G., Ed.; Elsevier Science: Amsterdam, The Netherlands, 1993; pp. 139–197.
4. Crognier, E. Biological adaptation and social behaviour. *Ann. Hum. Biol.* **2000**, *27*, 221–237. [CrossRef] [PubMed]
5. Gershman, S.J.; Horvitz, E.J.; Tenenbaum, J.B. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science* **2015**, *349*, 273–278. [CrossRef] [PubMed]
6. Hebb, D. *The Organization of Behaviour*; John Wiley & Sons: Hoboken, NJ, USA, 1949.
7. Minsky, M.; Papert, S. *Perceptrons. An Introduction to Computational Geometry*; MIT Press: Cambridge, MA, USA, 1969.
8. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef] [PubMed]
9. Haken, H. *Synergetic Computers and Cognition*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2004.
10. Hahn, T.; Nykvist, B. Are adaptations self-organized, autonomous, and harmonious? Assessing the social–ecological resilience literature. *Ecol. Soc.* **2017**, *22*, 12. [CrossRef]
11. Westley, F.R.; Tjornbo, O.; Schultz, L.; Olsson, P.; Folke, C.; Crona, B.; Bodin, O. A theory of transformative agency in linked social–ecological systems. *Ecol. Soc.* **2013**, *18*, 27. [CrossRef]
12. Deisboeck, T.S.; Berens, M.E.; Kansal, A.R.; Torquato, S.; Stemmer-Rachamimov, A.O.; Chiocca, E.A. Pattern of self-organization in tumour systems: Complex growth dynamics in a novel brain tumour spheroid model. *Cell Prolif.* **2001**, *34*, 115–134. [CrossRef]
13. Hassabis, D.; Kumaran, D.; Summerfield, C.; Botvinick, M. Neuroscience-Inspired Artificial Intelligence. *Neuron* **2017**, *95*, 245–258. [CrossRef]
14. Carpenter, G.A.; Grossberg, S. Discovering order in chaos: Stable self-organization of neural recognition codes. *Ann. N. Y. Acad. Sci.* **1987**, *504*, 33–51. [CrossRef]
15. Van Gerven, M. Computational Foundations of Natural Intelligence. *Front. Comput. Neurosci.* **2017**, *11*, 112. [CrossRef]
16. Grim, P. Computational Philosophy. Available online: https://plato.stanford.edu/entries/computational-philosophy/ (accessed on 29 April 2020).
17. Churchland, P.S.; Sejnowski, T. *The Computational Brain*; MIT Press: Cambridge, MA, USA, 1992.
18. Churchland, P.S. *Brain-Wise: Studies in Neurophilosophy*; MIT Press: Cambridge, MA, USA, 2002.
19. Lehmann, D.; Strik, W.K.; Henggeler, B.; Koenig, T.; Koukkou, M. Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking. I. Visual imagery and abstract thoughts. *Int. J. Psychophysiol.* **1998**, *29*, 1–11. [CrossRef]

20. Bassett, D.S.; Meyer-Lindenberg, A.; Achard, S.; Duke, T.; Bullmore, E. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19518–19523. [CrossRef] [PubMed]
21. Merzenich, M.M.; Nelson, R.J.; Stryker, M.P.; Cynader, M.S.; Schoppmann, A.; Zook, J.M. Somatosensory cortical map changes following digit amputation in adult monkeys. *J. Compar. Neurol.* **1984**, *224*, 591–605. [CrossRef] [PubMed]
22. Wall, J.T.; Xu, J.; Wang, X. Human brain plasticity: An emerging view of the multiple substrates and mechanisms that cause cortical changes and related sensory dysfunctions after injuries of sensory inputs from the body. *Brain Res. Rev.* **2002**, *39*, 181–215. [CrossRef]
23. Newman, M.E.J. Modularity and Community Structure in Networks. *Proc. Natl Acad. Sci. USA* **2006**, *103*, 8577. [CrossRef]
24. Tetzlaff, C.; Okujeni, S.; Egert, U.; Wörgötter, F.; Butz, M. Self-organized criticality in developing neuronal networks. *PLoS Comput. Biol.* **2010**, *6*, e1001013. [CrossRef]
25. Okujeni, S.; Egert, U. Self-organization of modular network architecture by activity-dependent neuronal migration and outgrowth. *eLife* **2019**, *8*, e47996. [CrossRef]
26. Kyan, M.; Muneesawang, P.; Jarrah, K.; Guan, L. Self-Organization. In *Unsupervised Learning*; Kyan, M., Muneesawang, P., Jarrah, K., Guan, L., Eds.; Wiley-IEEE Press: Hoboken, NJ, USA, 2014. [CrossRef]
27. Grossberg, S. Adaptive Resonance Theory. *Scholarpedia* **2013**, *8*, 1569. [CrossRef]
28. Kohonen, T. Physiological interpretation of the self-organizing map algorithm. *Neural Netw.* **1993**, *6*, 895–905. [CrossRef]
29. Kohonen, T. *Self-Organizing Maps*; Springer: Berlin/Heidelberg, Germany, 1995.
30. Berninger, B.; Bi, G.Q. Synaptic modification in neural circuits: A timely action. *BioEssays* **2002**, *24*, 212–222. [CrossRef]
31. Brette, R.; Rudolph, M.; Carnevale, T.; Hines, M.; Beeman, D.; Bower, J.M.; Diesmann, M.; Morrison, A.; Goodman, P.H.; Harris, F.C., Jr.; et al. Simulation of networks of spiking neurons: A review of tools and strategies. *J. Comput. Neurosci.* **2007**, *23*, 349–398. [CrossRef] [PubMed]
32. Delorme, A.; Thorpe, S.J. Spikenet: An event-driven simulation package for modelling large networks of spiking neurons. *Network Comput. Neural Sci.* **2003**, *14*, 613–627. [CrossRef]
33. Haider, B.; Schulz, D.P.A.; Häusser, M.; Carandini, M. Millisecond coupling of local field potentials to synaptic currents in the awake visual cortex. *Neuron* **2016**, *90*, 35–42. [CrossRef] [PubMed]
34. Bliss, T.V.; Cooke, S.F. Long-term potentiation and long-term depression: A clinical perspective. *Clinics* **2011**, *66*, 3–17. [CrossRef] [PubMed]
35. Cooke, S.F.; Bear, M.F. How the mechanisms of long-term synaptic potentiation and depression serve experience-dependent plasticity in primary visual cortex. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2013**, *369*, 20130284. [CrossRef]
36. Koch, H.; Garcia, A.J.; Ramirez, J.M. Network reconfiguration and neuronal plasticity in rhythm-generating networks. *Integr. Comp. Biol.* **2011**, *51*, 856–868. [CrossRef]
37. Motanis, H.; Seay, M.J.; Buonomano, D.V. Short-term synaptic plasticity as a mechanism for sensory timing. *Trends Neurosci.* **2018**, *41*, 701–711. [CrossRef]
38. Frank, M.G.; Cantera, R. Sleep, clocks, and synaptic plasticity. *Trends Neurosci.* **2014**, *37*, 491–501. [CrossRef]
39. Galuske, R.A.W.; Munk, M.H.J.; Singer, W. Relation between gamma oscillations and neuronal plasticity in the visual cortex. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 23317–23325. [CrossRef]
40. Lizbinski, K.M.; Dacks, A.M. Intrinsic and Extrinsic Neuromodulation of Olfactory Processing. *Front. Cell Neurosci.* **2018**, *11*, 424. [CrossRef]
41. Vecoven, N.; Ernst, D.; Wehenkel, A.; Drion, G. Introducing neuromodulation in deep neural networks to learn adaptive behaviours. *PLoS ONE* **2020**, *15*, e0227922. [CrossRef] [PubMed]
42. Grossberg, S. Acetylcholine Neuromodulation in Normal and Abnormal Learning and Memory: Vigilance Control in Waking, Sleep, Autism, Amnesia and Alzheimer's Disease. *Front. Neural Circuits* **2017**, *11*, 82. [CrossRef]
43. Darwin, C. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*; John Murray: London, UK, 1859.
44. Liu, Y. Natural Selection and Pangenesis: The Darwinian Synthesis of Evolution and Genetics. *Adv. Genet.* **2018**, *102*, 121–142. [PubMed]

45. Charlesworth, D.; Barton, N.H.; Charlesworth, B. The sources of adaptive variation. *Proc. Biol Sci.* **2017**, *284*, 1855. [CrossRef] [PubMed]
46. Grossberg, S. How hallucinations may arise from brain mechanisms of learning, attention, and volition. *J. Int. Neuropsychol. Soc.* **2000**, *6*, 579–588. [CrossRef]
47. Grossberg, S. Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2009**, *364*, 1223–1234. [CrossRef]
48. Joyce, K.E.; Hayasaka, S.; Laurienti, P.J. The Human Functional Brain Network Demonstrates Structural and Dynamical Resilience to Targeted Attack. *PLoS Comput. Biol.* **2013**, *9*, e1002885. [CrossRef]
49. Alstott, J.; Breakspear, M.; Hagmann, P.; Cammoun, L.; Sporns, O. Modeling the Impact of Lesions in the Human Brain. *PLoS Comput. Biol.* **2009**, *5*, e1000408. [CrossRef]
50. Maslov, S.; Sneppen, K. Specificity and Stability in Topology of Protein Networks. *Science* **2002**, *296*, 910–913. [CrossRef]
51. Silva, P.R.; Farias, T.; Cascio, F.; Dos Santos, L.; Peixoto, V.; Crespo, E.; Ayres, C.; Ayres, M.; Marinho, V.; Bastos, V.H.; et al. Neuroplasticity in visual impairments. *Neurol. Int.* **2018**, *10*, 7326. [CrossRef]
52. Bremner, J.D.; Elzinga, B.; Schmahl, C.; Vermetten, E. Structural and functional plasticity of the human brain in posttraumatic stress disorder. *Prog. Brain Res.* **2008**, *167*, 171–186.
53. Tanaka, J.; Horiike, Y.; Matsuzaki, M.; Miyazaki, T.; Ellis-Davies, G.C.; Kasai, H. Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **2008**, *319*, 1683–1687. [CrossRef] [PubMed]
54. Mermillod, M.; Bugaiska, A.; Bonin, P. The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects. *Front. Psychol.* **2013**, *4*, 504. [CrossRef]
55. Bramati, I.E.; Rodrigues, C.; Simões, E.L.; Melo, B.; Höfle, S.; Moll, J.; Lent, R.; Tovar-Moll, F. Lower limb amputees undergo long-distance plasticity in sensorimotor functional connectivity. *Sci. Rep.* **2019**, *9*, 2518. [CrossRef] [PubMed]
56. Sepulcre, J.; Liu, H.; Talukdar, T.; Martincorena, I.; Yeo, B.T.T.; Buckner, R.L. The Organization of Local and Distant Functional Connectivity in the Human Brain. *PLoS Comput. Biol.* **2010**, *6*, e1000808. [CrossRef]
57. Gogtay, N.; Giedd, J.N.; Lusk, L.; Hayashi, K.M.; Greenstein, D.; Vaituzis, A.C.; Nugent, T.F., 3rd; Herman, D.H.; Clasen, L.S.; Toga, A.W.; et al. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8174–8179. [CrossRef] [PubMed]
58. Fair, D.A.; Cohen, A.L.; Power, J.D.; Dosenbach, N.U.; Church, J.A.; Miezin, F.M.; Schlaggar, B.L.; Petersen, S.E. Functional brain networks develop from a "local to distributed" organization. *PLoS Comput. Biol.* **2009**, *5*, e1000381. [CrossRef] [PubMed]
59. Ungerleider, L.G.; Haxby, J.V. 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* **1994**, *4*, 157–165. [CrossRef]
60. Spillmann, L.; Dresp-Langley, B.; Tseng, C.H. Beyond the classic receptive field: The effect of contextual stimuli. *J. Vis.* **2015**. [CrossRef]
61. Bullmore, E.; Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **2009**, *10*, 186–198. [CrossRef]
62. Eguíluz, V.M.; Chialvo, D.R.; Cecchi, G.A.; Baliki, M.; Apkarian, A.V. Scale-free brain functional networks. *Phys. Rev. Lett.* **2005**, *94*, 018102. [CrossRef] [PubMed]
63. Hubel, D.H.; Wiesel, T.N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **1959**, *148*, 574–591. [CrossRef]
64. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154. [CrossRef] [PubMed]
65. Hubel, D.H.; Wiesel, T.N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) and of the cat. *J. Neurophysiol.* **1965**, *28*, 229–289. [CrossRef] [PubMed]
66. Hubel, D.H.; Wiesel, T.N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **1968**, *195*, 215–243. [CrossRef]
67. Li, H.H.; Chen, C.C. Surround modulation of global form perception. *J. Vis.* **2011**, *11*, 1–9. [CrossRef]
68. Muckli, L.; Vetter, P.; Smith, F. Predictive coding—Contextual processing in primary visual cortex V1. *J. Vis.* **2011**, *11*, 25. [CrossRef]
69. Muckli, L.; Petro, L.S. Network interactions: Non-geniculate input to V1. *Curr. Opin. Neurobiol.* **2013**, *23*, 195–201. [CrossRef]

70. Grossberg, S.; Swaminathan, G. A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: Development, attention and bistability. *Vis. Res.* **2004**, *44*, 1147–1187. [CrossRef]
71. Grossberg, S.; Yazdanbakhsh, A. Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vis. Res.* **2005**, *45*, 1725–1743. [CrossRef]
72. Dresp-Langley, B.; Grossberg, S. Neural Computation of Surface Border Ownership and Relative Surface Depth from Ambiguous Contrast Inputs. *Front. Psychol.* **2016**, *7*, 1102. [CrossRef] [PubMed]
73. Dresp-Langley, B.; Reeves, A.; Grossberg, S. Editorial: Perceptual Grouping—The State of The Art. *Front. Pscychol.* **2017**, *8*, 67. [CrossRef] [PubMed]
74. Dresp-Langley, B. Bilateral Symmetry Strengthens the Perceptual Salience of Figure against Ground. *Symmetry* **2019**, *11*, 225. [CrossRef]
75. Dresp-Langley, B.; Monfouga, M. Combining Visual Contrast Information with Sound Can Produce Faster Decisions. *Information* **2019**, *10*, 346. [CrossRef]
76. Grossberg, S. The link between brain learning, attention, and consciousness. *Conscious. Cognit.* **1999**, *8*, 1–44. [CrossRef] [PubMed]
77. Grossberg, S.; Myers, C.W. The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychol. Rev.* **2000**, *107*, 735–767. [CrossRef]
78. Helekar, S.A. On the possibility of universal neural coding of subjective experience. *Conscious. Cognit.* **1999**, *8*, 423–446. [CrossRef]
79. Dresp-Langley, B.; Durup, J. A plastic temporal brain code for conscious state generation. *Neural Plast.* **2009**, *2009*, 482696. [CrossRef]
80. Hoffmann, H.; Payton, D.W. Optimization by Self-Organized Criticality. *Sci. Rep.* **2018**, *8*, 2358. [CrossRef]
81. Singer, W. The brain as a self-organizing system. *Eur. Arch. Psychiatr. Neurol. Sci.* **1986**, *236*, 4–9. [CrossRef]
82. Wilson, S.; Moore, C. S1 somatotopic maps. *Scholarpedia* **2015**, *10*, 8574. [CrossRef]
83. Braun, C.; Heinz, U.; Schweizer, R.; Wiech, K.; Birbaumer, N.; Topka, H. Dynamic organization of the somatosensory cortex induced by motor activity. *Brain* **2001**, *124*, 2259–2267. [CrossRef] [PubMed]
84. Arber, S. Motor circuits in action: Specification, connectivity, and function. *Neuron* **2012**, *74*, 975–989. [CrossRef] [PubMed]
85. Tripodi, M.; Arber, S. Regulation of motor circuit assembly by spatial and temporal mechanisms. *Curr. Opin. Neurobiol.* **2012**, *22*, 615–623. [CrossRef]
86. Weiss, T.; Miltner, W.H.R.; Huonker, R.; Friedel, R.; Schmidt, I.; Taub, E. Rapid functional plasticity of the somatosensory cortex after finger amputation. *Exp. Brain Res.* **2000**, *134*, 199–203. [CrossRef]
87. Young, R.W. Evolution of the human hand: The role of throwing and clubbing. *J. Anat.* **2003**, *202*, 165–174. [CrossRef]
88. Kinoshita, H.; Kawai, S.; Ikuta, K. Contributions and co-ordination of individual fingers in multiple finger prehension. *Ergonomics* **1995**, *38*, 1212–1230. [CrossRef]
89. Latash, M.L.; Zatsiorsky, V.M. Multi-finger prehension: Control of a redundant mechanical system. *Adv. Exp. Med. Biol.* **2009**, *629*, 597–618.
90. Oku, T.; Furuya, S. Skilful force control in expert pianists. *Exp. Brain Res.* **2017**, *235*, 1603–1615. [CrossRef]
91. Zatsiorsky, V.M.; Latash, M.L. Multifinger prehension: An overview. *J. Mot. Behav.* **2008**, *40*, 446–476. [CrossRef]
92. Sun, Y.; Park, J.; Zatsiorsky, V.M.; Latash, M.L. Prehension synergies during smooth changes of the external torque. *Exp. Brain Res.* **2011**, *213*, 493–506. [CrossRef] [PubMed]
93. Wu, Y.H.; Zatsiorsky, V.M.; Latash, M.L. Static prehension of a horizontally oriented object in three dimensions. *Exp. Brain Res.* **2002**, *216*, 249–261. [CrossRef] [PubMed]
94. Cha, S.M.; Shin, H.D.; Kim, K.C.; Park, J.W. Comparison of grip strength among six grip methods. *J. Hand Surg Am.* **2014**, *39*, 2277–2284. [CrossRef] [PubMed]
95. Cai, A.; Pingel, I.; Lorz, D.; Beier, J.P.; Horch, R.E.; Arkudas, A. Force distribution of a cylindrical grip differs between dominant and nondominant hand in healthy subjects. *Arch. Orthop. Trauma Surg.* **2018**, *138*, 1323–1331. [CrossRef] [PubMed]
96. Bohannon, R.W. Grip strength: A summary of studies comparing dominant and non-dominant limb measurements. *Percept. Mot. Skills* **2003**, *96*, 728–730. [CrossRef]
97. Johansson, R.S.; Cole, K.J. Sensory-motor coordination during grasping and manipulative actions. *Curr. Opin. Neurobiol.* **1992**, *2*, 815–823. [CrossRef]

98. Eliasson, A.C.; Forssberg, H.; Ikuta, K.; Apel, I.; Westling, G.; Johansson, R. Development of human precision grip V. Anticipatory and triggered grip actions during sudden loading. *Exp. Brain Res.* **1995**, *106*, 425–433.
99. Jenmalm, P.; Johansson, R.S. Visual and somatosensory information about object shape control manipulative fingertip forces. *J. Neurosci.* **1997**, *17*, 4486–4499. [CrossRef]
100. Li, K.W.; Yu, R. Assessment of grip force and subjective hand force exertion under handedness and postural conditions. *Appl. Ergon.* **2011**, *42*, 929–933. [CrossRef]
101. Aravena, P.; Delevoye-Turrell, Y.; Deprez, V.; Cheylus, A.; Paulignan, Y.; Frak, V.; Nazir, T. Grip force reveals the context sensitivity of language-induced motor activity during "action words" processing: Evidence from sentential negation. *PLoS ONE* **2012**, *7*, e50287. [CrossRef]
102. González, A.G.; Rodríguez, D.R.; Sanz-Calcedo, J.G. Ergonomic analysis of the dimension of a precision tool handle: A case study. *Procedia Manuf.* **2017**, *13*, 1336–1343. [CrossRef]
103. De Mathelin, M.; Nageotte, F.; Zanne, P.; Dresp-Langley, B. Sensors for Expert Grip Force Profiling: Towards Benchmarking Manual Control of a Robotic Device for Surgical Tool Movements. *Sensors* **2019**, *19*, 4575. [CrossRef] [PubMed]
104. Batmaz, A.U.; Falek, A.M.; Zorn, L.; Nageotte, F.; Zanne, P.; de Mathelin, M.; Dresp-Langley, B. Novice and expert behavior while using a robot controlled surgery system. In Proceedings of the 13th IASTED International Conference on Biomedical Engineering (BioMed), Innsbruck, Austria, 20–21 February 2017.
105. Batmaz, A.U.; Falek, M.A.; de Mathelin, M.; Dresp-Langley, B. Tactile sensors for measuring effects of sight, movement, and sound on handgrip forces during hand-tool interaction. *Preprints* **2017**. [CrossRef]
106. Batmaz, A.U.; de Mathelin, M.; Dresp-Langley, B. Seeing virtual while acting real: Visual display and strategy effects on the time and precision of eye-hand coordination. *PLoS ONE* **2017**, *12*, e0183789. [CrossRef]
107. Kawamura, S.; Svinin, M. *Advances in Robot Control: From Everyday Physics to Human-Like Movements*; Springer: New York, NY, USA, 2006.
108. Dresp-Langley, B. Towards Expert-Based Speed–Precision Control in Early Simulator Training for Novice Surgeons. *Information* **2018**, *9*, 316. [CrossRef]
109. Dresp-Langley, B. Why the brain knows more than we do: Non-conscious representations and their role in the construction of conscious experience. *Brain Sci.* **2011**, *2*, 1–21. [CrossRef]
110. Feigl, H. The "Mental" and the "Physical". Available online: https://conservancy.umn.edu/handle/11299/184614 (accessed on 7 May 2020).
111. Dresp-Langley, B.; Wandeto, J.M.; Nyongesa, H.K.O. Using the Quantization Error from Self-Organizing Map Output for Fast Detection of Critical Variations in Image Time Series. Available online: https://www.openscience.fr/Donnees-image-et-decision-detection-automatique-de-variations-dans-des-series (accessed on 7 May 2020).
112. Wandeto, J.M.; Dresp-Langley, B. The quantization error in a Self-Organizing Map as a contrast and color specific indicator of single-pixel change in large random patterns. *Neural Netw.* **2019**, *119*, 273–285. [CrossRef]
113. Le Cun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *215*, 437.
114. Dresp-Langley, B.; Ekseth, O.K.; Fesl, J.; Gohshi, S.; Kurz, M.; Sehring, H.W. Occam's Razor for *Big Data*? On Detecting Quality in Large Unstructured Datasets. *Appl. Sci.* **2019**, *9*, 3065. [CrossRef]
115. Vahdat, S.; Darainy, M.; Thiel, A.; Ostry, D.J. A Single Session of Robot-Controlled Proprioceptive Training Modulates Functional Connectivity of Sensory Motor Networks and Improves Reaching Accuracy in Chronic Stroke. *Neurorehabil. Neural Repair* **2019**, *33*, 70–81. [CrossRef]
116. Miall, R.C.; Kitchen, N.M.; Nam, S.H.; Lefumat, H.; Renault, A.G.; Orstavik, K.; Cole, J.D.; Sarlegna, F.R. Proprioceptive loss and the perception, control and learning of arm movements in humans: Evidence from sensory neuronopathy. *Exp. Brain Res.* **2018**, *236*, 2137–2155. [CrossRef] [PubMed]
117. Elangovan, N.; Yeh, I.L.; Holst-Wolf, J.; Konczak, J. A robot-assisted sensorimotor training program can improve proprioception and motor function in stroke survivors. In Proceedings of the 16th International Conference on Rehabilitation Robotics, Toronto, ON, Canada, 24–28 June 2019.



big data and cognitive computing

MDPI

Article

#*lockdown*: Network-Enhanced Emotional Profiling in the Time of COVID-19

Massimo Stella [1,*] and Valerio Restocchi [2] and Simon De Deyne [3]

[1] Complex Science Consulting, 73100 Lecce, Italy
[2] School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK; v.restocchi@ed.ac.uk
[3] Human Complex Data Hub, School of Psychological Sciences, University of Melbourne, Melbourne, VIC 3010, Australia; simon.dedeyne@unimelb.edu.au
* Correspondence: massimo.stella@inbox.com

Received: 8 May 2020; Accepted: 6 June 2020; Published: 16 June 2020

Abstract: The COVID-19 pandemic forced countries all over the world to take unprecedented measures, like nationwide lockdowns. To adequately understand the emotional and social repercussions, a large-scale reconstruction of how people perceived these unexpected events is necessary but currently missing. We address this gap through social media by introducing MERCURIAL (Multi-layer Co-occurrence Networks for Emotional Profiling), a framework which exploits linguistic networks of words and hashtags to reconstruct social discourse describing real-world events. We use MERCURIAL to analyse 101,767 tweets from Italy, the first country to react to the COVID-19 threat with a nationwide lockdown. The data were collected between the 11th and 17th March, immediately after the announcement of the Italian lockdown and the WHO declaring COVID-19 a pandemic. Our analysis provides unique insights into the psychological burden of this crisis, focussing on—(i) the Italian official campaign for self-quarantine (*#iorestoacasa*), (ii) national lockdown (*#italylockdown*), and (iii) social denounce (*#sciacalli*). Our exploration unveils the emergence of complex emotional profiles, where anger and fear (towards political debates and socio-economic repercussions) coexisted with trust, solidarity, and hope (related to the institutions and local communities). We discuss our findings in relation to mental well-being issues and coping mechanisms, like instigation to violence, grieving, and solidarity. We argue that our framework represents an innovative thermometer of emotional status, a powerful tool for policy makers to quickly gauge feelings in massive audiences and devise appropriate responses based on cognitive data.

Keywords: COVID-19; social media; hashtag networks; emotional profiling; cognitive science; network science; sentiment analysis; computational social science

1. Introduction

The stunningly quick spread of the COVID-19 pandemic catalysed the attention of worldwide audiences, overwhelming individuals with a deluge of often contrasting content about the severity of the disease, the uncertainty of its transmission mechanisms, and the asperity of the measures taken by most countries to fight it [1–4]. Although these policies have been seen as necessary, they had a tremendous impact on the mental well-being of large populations [5] for a number of reasons. Due to lockdowns, many are facing financial uncertainty, having lost or being on the verge of losing their source of income. Moreover, there is much concern about the disease itself, and most people fear for their own health and that of their loved ones [6], further fueled by *infodemics* [1–3]. Finally, additional distress is caused by the inability of maintaining a normal life [7]. The extent of the impact of these factors is such that, in countries greatly struck by COVID-19 such as China, the population started to develop symptoms of post-traumatic stress disorder [8].

During this time more than ever, people have shared their emotions on social media. These platforms provide an excellent emotional thermometer of the population, and have been widely explored in previous studies investigating how online social dynamics promote or hamper content diffusion [1,2,9–11] and the adoption of specific positive/negative attitudes and behaviours [9,12,13].

1.1. Research Aim

Building on the above evidence, our goal is to draw a comprehensive quantitative picture of people's emotional profiles, emerging during the COVID-19 crisis, through a cognitive analysis of online social discourse. We achieve this by introducing MERCURIAL (Multi-layer Co-occurrence Networks for Emotional Profiling), a framework that combines cognitive network science [14–16] with computational social sciences [2,9,12,17,18]. Before outlining the methods and main contributions of our approach, we briefly review existing research on understanding emotions in social media.

1.2. Past Approaches Bridging Cognitive, Computer and Network Science

Much of the research on emotions in social media has been consolidated into two themes. On the one hand, there is the data science approach, which mostly focused over large-scale positive/negative sentiment detection [9]. A prominent example is the Hedonometer [19], a multi-language tool measuring over time the positive sentiment of Twitter discourse based on word frequency. Recent approaches identified the relevance of tracing more complex affect patterns for understanding social dynamics [16,20–22]. On the other hand, cognitive science research makes use of small-scale analysis tools, but explores the observed phenomena in much more detail in the light of its theoretical foundations [23–25]. Specifically, in cognitive science the massive spread of semantic and emotional information through verbal communication represent long-studied phenomena, known as *cognitive contagion* [25] and *emotional contagion* [24–26], respectively. This research suggests that ideas are composed of a cognitive component and an emotional content, much alike viruses containing the genomic information necessary for their replication [1]. Both these types of contagion happen when an individual is affected in their behaviour by an idea. Emotions elicited by ideas can influence users' behaviour without their awareness, resulting in the emergence of specific behavioural patterns such as implicit biases [24]. Unlike pathogen transmission, no direct contact is necessary for cognitive and emotional contagion to take place, since both are driven by information processing and diffusion, like it happens through social media [27,28]. In particular, during large-scale events, ripples of emotions can rapidly spread across information systems [28] and have dramatic effects, as it has recently been demonstrated in elections and social movements [12,26,29].

At the intersection of data- and cognitive science is emotional profiling, a set of techniques which enables the reconstruction of how concepts are emotionally perceived and assembled in user-generated content [9,15,17–20,30]. Emotional profiling conveys information about basic affective dimensions such how positive/negative or how arousing a message is, and also includes the analysis of more fine-grained emotions such as *fear* or *trust* that might be associated with the lockdown and people's hopes for the future [9,23,31].

Recently, an emerging important line of research has shown that reconstructing the knowledge embedded in messages through social and information network models [14,32,33] successfully highlight important phenomena in a number of contexts, ranging from the diffusion of hate speech during massive voting events [12] to reconstructing personality traits from social media [17]. Importantly, to reconstruct knowledge embedded in tweets, recent work has successfully merged data science and cognitive science, introducing linguistic networks of co-occurrence relationships between words in sentences [16,22,34] and between hashtags in tweets [12]. However, an important shortfall of these works is that these two types of networked knowledge representations were not merged together, thus missing on the important information revealed by studying their interdependence.

1.3. Main Contributions

We identify three important contributions that distinguish our paper from previous literature, and make a further step towards consolidating *cognitive network science* [14] as a paradigm suitable to analyse people's emotions.

First, we introduce a new framework exploiting the interdependence between hashtags and words, addressing the gap previously discussed. This framework, multi-layer co-occurrence networks for emotional profiling (MERCURIAL), combines both the semantic structure encoded through the co-occurrence of hashtags and the textual message to construct a multi-layer lexical network [35]. The multilayer network represented by MERCURIAL is a network of networks, where there is one hashtag co-occurrence network and many word co-occurrence networks. Each hashtag co-occurrence link identifies a word co-occurrence network, indicating language structure in tweets with certain hashtag combinations. This multi-layer network structure allows us to contextualise hashtags and, therefore, improve the analysis of their meaning. Importantly, these networks can be used to identify which concepts or words contribute to different emotions and how central they are.

Second, in contrast to previous work, which largely revolved around English tweets [4,21], the current study focusses on Italian Twitter messages. There are several reasons why the emotional response of Italians is particularly interesting. Specifically, (i) Italy was the first Western country to experience a vast number of COVID-19 clusters; (ii) the Italian government was the first to declare a national lockdown (cf. https://en.wikipedia.org/wiki/COVID-19_pandemic_lockdown_in_Italy, Last Access: 10 June 2020); (iii), the Italian lockdown was announced on 10th March, one day before the World Health Organization (WHO) declared the pandemic status of COVID-19. This enables us to address the urgent need of measuring the emotional perceptions and reactions to social distancing, lockdown, and, more generally, the COVID-19 pandemic.

Third, thanks to MERCURIAL, we obtain richer and more complex emotional profiles that we analyse through the lens of established psychological theories of emotion. This is a fundamental step in going beyond positive/neutral/negative sentiment and to provide accurate insights on the mental well-being of a population. To this end, we take into account three of the most trending hashtags, *#iorestoacasa* (English: "I stay at home"), *#sciacalli* (English: "jackals"), and *#italylockdown*, as representative of positive, negative, and neutral social discourse, respectively. We use these hashtags as a starting point to build multi-layer networks of word and hashtag co-occurrence, from which we derive our profiles. Our results depict a complex map of emotions, suggesting that there is co-existence and polarisation of conflicting emotional states, importantly fear and trust towards the lockdown and social distancing. The combination of these emotions, further explored through semantic network analysis, indicates mournful submission and acceptance towards the lockdown, perceived as a measure for preventing contagion but with negative implications over economy. As further evidence of the complexity of the emotional response to the crisis, we also find strong signals of hope and social bonding, mainly in relation to social flash mobs, and interpreted here as psychological responses to deal with the distress caused by the threat of the pandemic.

1.4. Manuscript Outline

The paper is organised as follows. In the Methods section we describe the data we used to perform our analysis, and describe MERCURIAL in detail. In Section 3 we present the emotional profiles obtained from our data, which are then discussed in more detail in the section Section 4. Finally, the last section highlights the psychological implications of our exploratory investigation and its potential for follow-up monitoring of COVID-19 perceptions in synergy with other datasets/approaches.

We argue that our findings represent an important first step towards monitoring both mental well-being and emotional responses in real time, offering policy-makers a framework to make timely data-informed decisions.

2. Methods

In this section we describe the methodology employed to collect our data and perform the emotional profiling analysis. First, we describe the dataset and how it was retrieved. Then, we introduce co-occurrence networks, and specifically our novel method that combines hasthag co-occurrence with word co-occurrence on multi-layer networks. Finally, we describe the cognitive science framework we used to perform the emotional profiling analysis on the so-obtained networks.

2.1. Data

We gathered 101,767 tweets in Italian to monitor how online users perceived the COVID-19 pandemic and its repercussions in Italy. These tweets were gathered by crawling messages containing three trending hashtags of relevance for the COVID-19 outbreak in Italy and expressing three different sentiment polarities:

- *#iorestoacasa* (English: "I stay at home"), a positive-sentiment hashtag introduced by the Italian Government in order to promote a responsible attitude during the lockdown;
- *#sciacalli*, (English: "jackals"), a negative sentiment hashtag used by online users in order to address unfair behaviour rising during the health emergency;
- *#italylockdown*, a neutral sentiment hashtag indicating the application of lockdown measures all over Italy.

We refer to *#iorestoacasa*, *#sciacalli* and *#italylockdown* as *focal hashtags* to distinguish them from other hashtags. We collected the tweets through *Complex Science Consulting* (*@ComplexConsult*), which was authorised by Twitter, and used the *ServiceConnect* crawler implemented in Mathematica 11.3. The collection of tweets comprises 39,943 tweets for *#iorestoacasa*, 26,999 for *#sciacalli* and 34,825 for *#italylockdown*. Retweets of the same text message were not considered. For each tweet, the language was detected. Pictures, links, and non-Italian content was discarded and stop-words (i.e., words without intrinsic meaning such as "di" (English: "of") and "ma" (English: "but") removed. Other interesting datasets with tweets about COVID-19 are available in References [21,36].

2.2. Multi-Layer Co-Occurrence Networks

Word co-occurrence networks have been successfully used to characterise a wide variety of phenomena related to language acquisition and processing [16,37,38]. Recently, researchers have also used hashtags to investigate various aspects of social discourse. For instance, Stella et al. [12] showed that hashtag co-occurrence networks were able to characterise important differences in the social discourses promoted by opposing social groups during the Catalan referendum. In this work we introduce MERCURIAL (Multi-layer Co-occurrence Networks for Emotional Profiling), a framework combining:

- Hashtag co-occurrence networks (or hashtag networks) [12]. Nodes represent hashtags and links indicate the co-occurrence of any two nodes in the same tweet.
- Word co-occurrence networks (or word networks) [16]. Nodes represent words and links represent the co-occurrence of any two words one after the other in a tweet without stop-words (i.e., words without an intrinsic meaning).

We combine these two types of networks in a multi-layer network to exploit the interdependence between hashtags and words. This new, resulting network enables us to contextualise hashtags, and capture their real meaning through context, thereby enhancing the accuracy of the emerging emotional profile. To build the multi-layer network, we first build the single hashtag and word layers.

For sake of simplicity, word networks are unweighted and undirected. Robustness checks were performed for various cut-offs in co-occurrence frequency, for example, pruning co-occurrence networks according to co-occurrence weights did not change the most central hashtags or global emotional profiles. Note that the hashtag network was kept at a distinct level from word networks, for example, common words were not explicitly linked with hashtags. As reported in Figure 1,

each co-occurrence link between any two hashtags A and B (*#coronavirus* and *#restiamoacasa* in the figure) is relative to a word network, including all words co-occurring in all tweets featuring hashtags A and B.

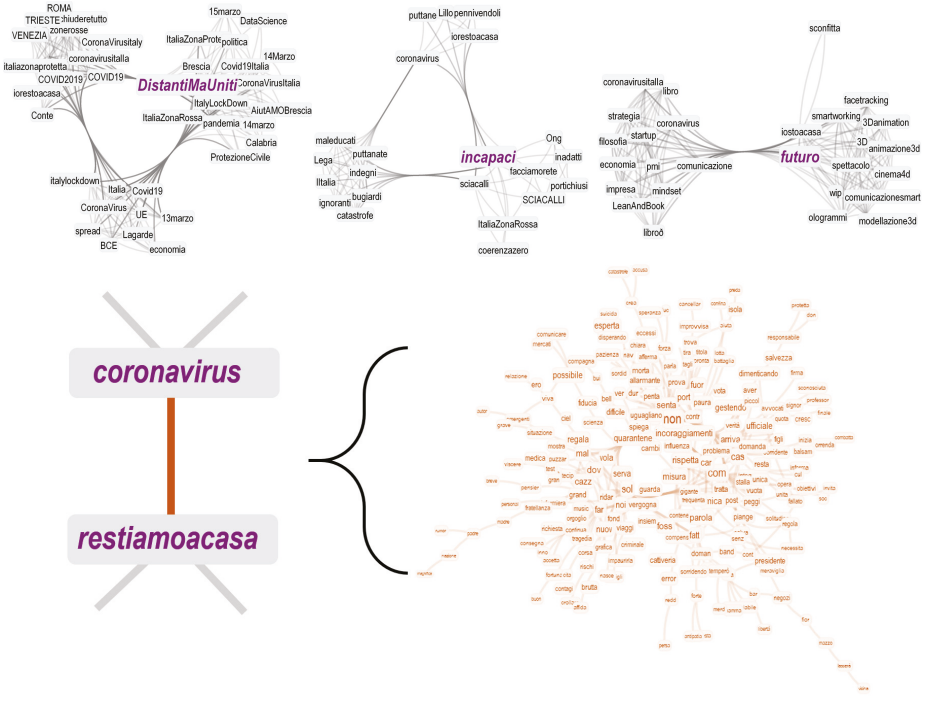


Figure 1. Top: Example of co-occurrence networks for different hashtags: *#distantimauniti* (English: distant but united) in *#iorestoacasa* on the left, *#incapaci* (English: inept) in *#sciacalli* in the middle, and *#futuro* (English: future) in *#italylockdown* on the right. Clusters of co-occurring hashtags were obtained through spectral clustering [39]. These clusters highlight the co-occurrence of European-focused content, featuring hashtags like *#BCE* (i.e., European Central Bank), *#Lagarde* and *#spread* (i.e., spread between Italian and German bonds) together with social distance practices related to *#iorestoacasa*. **Bottom:** In MERCURIAL, any link in a co-occurrence network of hashtags (left) corresponds to a collection of tweets whose words co-occur according to a word network (right). Larger words have a higher closeness centrality.

The hashtag and word networks capture the co-occurrence of lexical entities within the structured online social discourse. Words possess meaning in language [32] and their network assembly is evidently a linguistic network. Similar to words in natural language, hashtags possess linguistic features that express a specific meaning and convey rich affect patterns [12]. The resulting networks capture the meaning of a collection of tweets by identifying which words/hashtags co-occurred together.

This knowledge embedded in hashtag networks was used in order to identify the most relevant or central terms associated within a given collection of thematic tweets. Rather than using frequency to indicate centrality, which makes it difficult to compare hashtags that do not co-occur in the same message, the current work relies on distance-based measures to detect how central a hashtag is in the network.

The first measure that implements this notion is closeness centrality. Closeness $c(i)$ identifies how many links connect i to all its N neighbours and is formalised as follows:

$$c(i) = \frac{N}{\sum_{j=1}^{N} d_{ij}}, \quad (1)$$

where d_{ij} is the network distance between i and j, that is, the smallest amount of links connecting nodes i and j. In co-occurrence networks, nodes (i.e., hashtags and words) with a higher closeness tend to co-occur more often with each other or with other relevant nodes at short network distance. We expect that rankings of closeness centrality will reveal the most central hashtags in the networks for *#iorestoacasa*, *#sciacalli* and *#italylockdown*, in line with previous work in which closeness centrality was used to measure language acquisition and processing [14,33,40].

Importantly, closeness is a more comprehensive approach compared to the simpler frequency analysis. Imagine a collection of hashtags $A, B, C, D, \ldots$. Computing the frequency of hashtag A co-occurring with hashtag B is informative about the frequency of the so-called 2-grams "AB" or "BA" but it does not consider how those hashtags co-occur with C, D, and so forth. In other words, a 2-gram captures the co-occurrence of two specific hashtags within tweets but does not provide the simultaneous structure of co-occurrences of all hashtags across tweets, for which a network of pairwise co-occurrences is required. On such a network, closeness can then highlight hashtags at short distance from all others, that is, co-occurring in a number of contexts in the featured discourse.

In addition to closeness, we also use *graph distance entropy* to measure centrality. This centrality measure captures which hashtags are uniformly closer to all other hashtags in a connected network. Combining closeness with graph distance entropy led to successfully identifying words of relevance in conceptual networks with a few hundreds of nodes [33].

The main idea behind graph distance entropy is that it provides info about the spread of the distribution of network distances between nodes (i.e., shortest path), a statistical quantity that cannot be extracted from closeness (which is, conversely, a mean inverse distance). Considering the set $\mathbf{d}^{(i)} \equiv (d_{i1}, \ldots, d_{ij}, \ldots, d_{iN})$ of distances between i and any other node j connected to it ($1 \leq j \leq N$) and $M_i = Max(\mathbf{d}^{(i)})$, then graph distance entropy is defined as:

$$h(i) = -\frac{1}{\log(M_i - 1)} \sum_{k=1}^{M_i - 1} p_k^{(i)} \log p_k^{(i)}, \quad (2)$$

where p_k is the probability of finding a distance equal to k. Therefore, $h(i)$ is a Shannon entropy of distances and it ranges between 0 and 1. In general, the lower the entropy, the more a node resembles a star centre [35] and is at equal distances from all other nodes. Thus, nodes with a lower $h(i)$ and a higher closeness are more uniformly close to all other connected nodes in a network. Words with *simultaneously* low graph distance entropy and high closeness were found to be prominent words for early word learning [35] and mindset characterisation [33].

2.3. Attributing Meaning and Emotions to Focal Hashtags by Using Word Networks

In addition to hashtag networks, we also build word networks obtained from a collection of tweets containing any combination of the focal hashtags *#iorestocasa* or *#sciacalli* and *#coronavirus*. For all tweets containing a given set of hashtags, we performed the following:

1. Subdivide the tweet in sentences and delete all stop-words from each sentence, preserving the original ordering of words;
2. Stem all the remaining words, that is, identify the root or stem composing a given word. In a language such as Italian, in which there is a number of ways of adding suffixes to words, word stemming is essential in order to recognise the same word even when it is inflected for different gender, number or as a verb tense. For instance, *abbandoneremo* (we will abandon) and *abbandono* (abandon, abandonment) both represent the same stem *abband*;
3. Draw links between a stemmed word and its subsequent one. Store the resulting edge list of word co-occurrences.

4. Sentences containing a negation (i.e., "not") underwent an additional step parsing their syntactic structure. This was done in order to identify the target of negation (e.g., in "this is not peace", the negation refers to "peace"). Syntactic dependencies were not used for network construction but intervened in emotional profiling, instead (see below).

The resulting word network also captures syntactic dependencies between words [16] related by online users to a specific hashtag or combination of hashtag. We used closeness centrality to detect the relevance of words for a given hashtag. Text pre-processing such as word stemming was performed with the R package Snowball C while syntactic dependencies was performed using Mathematica 11.3, which was also used to extract networks and compute network metrics.

The presence of hashtags in word networks provided a way of linking words, which express common language, with hashtags, which express content but also summarise the topic of a tweet. Consequently, by using this new approach, the meaning attributed by users to hashtags can be inferred not only from hashtag co-occurrence but also from word networks. An example of MERCURIAL, featuring hashtag-hashtag and word-word co-occurrences, is reported in Figure 1 (bottom). In this example, hashtags *#coronavirus* and *#restiamoacasa* co-occurred together (left) in tweets featuring many co-occurring words (right). The resulting word network shows relevant concepts such as "incoraggiamenti" (English: encouragement) and "problemi" (English: problems), highlighting a positive attitude towards facing problems related to the pandemic. More in general, the attribution and reconstruction of such meaning was explored by considering conceptual relevance and emotional profiling in one or several word networks related to a given region of a hashtag co-occurrence network.

2.4. Emotional Profiling

As a first data source for emotional profiling, this work also used valence and arousal data from Warriner and colleagues [41], whose combination can reconstruct emotional states according to the well-studied circumplex model of affect [31,42]. In psycholinguistics, word valence expresses how positively/negatively a concept is perceived (equivalently to sentiment in computer science). The second dimension, arousal, indicates the alertness or lethargy inspired by a concept. Having a high arousal and valence indicates excitement and joy, whereas a negative valence combined with a high arousal can result in anxiety and alarm [31]. Finally, some studies also include dominance or potency as a measure of the degree of control experienced [41]. However, for reasons of conciseness, we focus on the two primary dimensions of affect: valence and arousal.

Going beyond the standard positive/negative/neutral sentiment intensity is of utmost importance for characterising the overall online perception of massive events [9]. Beyond the primary affective dimension of sentiment, the affect associated with current events [12] can also be described in terms of arousal [43] and of basic emotions such as fear, disgust, anger, trust, joy, surprise, sadness, and anticipation. These emotions represent basic building blocks of many complex emotional states [24], and they are all self-explanatory except for anticipation, which indicates a projection into future events [18]. Whereas fear, disgust, and anger (trust and joy) elicit negative (positive) feedback, surprise, sadness and anticipation have been recently evaluated as neutral emotions, including both positive and negative feedback reactions to events in the external world [44].

To attribute emotions to individual words, we use the National Research Council of Canada (NRC) lexicon [18] and the circumplex model [31]. These two approaches allow us to quantify the *emotional profile* of a set of words related to hashtags or combinations of hashtags. The NRC lexicon enlists words eliciting a given emotion. The circumplex model attributes valence and arousal scores to words, which in turn determine their closest emotional states. Because datasets of similar size were not available for Italian, the data from the NRC lexicon and the Warriner norms were translated from English to Italian using a forward consensus translation of Google Translate, Microsoft Bing and DeepL translator, which was successfully used in previous investigations with Italian [45]. Although the valence of some concepts might change across languages [41], word stemming related several scores to the same stem, for example, scores for "studio" (English: "study") and "studiare"

(English: "to study") were averaged together and the average attributed to the stem root "stud". In this way, even if non-systematic cross-language valence shifting introduced inaccuracy in the score for one word (e.g., "studiare"), averaging over other words relative to the same stem reduced the influence of such inaccuracy. No statistically significant difference ($\alpha = 0.05$) was found between the emotional profiles of 200 Italian tweets, including 896 different stems, and their automatic translations in English, holding for each dimension separately (z-scores < 1.96).

Then, we build emotional profiles by considering the distribution of words eliciting a given emotion/valence/arousal and associated to specific hashtags in tweets. Assertive tweets with no negation were evaluated directly through a bag of words model, that is, by directly considering the words composing them. Tweets including negations underwent an additional intermediate step where words syntactically linked to the negation were substituted with their antonyms [46] and then evaluated. Source-target syntactic dependencies were computed in Mathematica 11.3 and all words targeted by a negation word (i.e., *no*, *non* and *nessuno* in Italian) underwent the substitution with their antonyms.

To determine whether the observed emotional intensity $r(i)$ of a given emotion in a set S of words was compatible with random expectation, we perform a statistical test (Z-test) using the NRC dataset. Remember that emotional intensity here was measured in terms of richness or count of words eliciting a given emotion in a given network. As a null model, we use random samples as follows: let us denote by m the number of words stemmed from S that are also in the NRC dataset. Then, m words from the NRC lexicon are sampled uniformly at random and their emotional profile is compared against that of the empirical sample. We repeated this random sampling 1000 times for each single empirical observed emotional profile $\{r(i)\}_i$. To ensure the resulting profiles are indeed compatible with a Gaussian distribution, we performed a Kolmogorov-Smirnov test ($\alpha = 0.05$). All the tests we performed gave random distributions of emotional intensities compatible with a Gaussian distribution, characterised by a mean random intensity for emotion i, $r^*(i)$ and a standard deviation $\sigma^*(i)$. For each emotion, a z-score was computed:

$$z_i = \frac{r(i) - r^*(i)}{\sigma^*(i)}. \tag{3}$$

In the remainder of the manuscript, every emotional profile incompatible with random expectation was highlighted in black or marked with a check. Since we used a two-tailed Z-test (with a significance level of 0.05), this means that an emotional richness can either be higher or lower than random expectation. Notice that emotional richness and random expectations are different across different emotions, since in the NRC lexicon there are more words eliciting fear and trust and fewer words eliciting other emotions like joy or surprise. For this reason, it is important to consider z-scores for each individual emotion.

3. Results

The investigated corpus of tweets represents a complex multilevel system, where conceptual knowledge and emotional perceptions are entwined on a number of levels. Tweets are made of text and include words, which convey meaning [32]. From the analysis of word networks, we can obtain information on the organisation of knowledge proper of social media users, which is embedded in their generated content [16]. However, tweets also convey meaning through the use of hashtags, which can either refer to specific words or point to the overall topic of the whole tweet. Both words and hashtags can evoke emotions in different contexts, thus giving rise to complex patterns [17]. Similar to words in natural language, the same hashtags can be perceived and used in language differently by different users, according to the context.

The simultaneous presence of word- and hashtag-occurrences in tweets is representative of the knowledge shared by social media users when conveying specific content and ideas.

This interconnected representation of knowledge can be exploited by simultaneously considering both hashtag-level and word-level information, since words specify the meaning attributed to hashtags.

In this section we use MERCURIAL to analyse the data collected. We do so by characterising the hashtag networks, both in terms of meaning and emotional profiles. Precedence is given to hashtags as they not only convey meaning as individual linguistic units but also represent more general-level topics characterising the online discourse. Then, we inter-relate hashtag networks with word networks. Finally, we perform the emotional profiling of hashtags in specific contexts. The combination of word- and hashtag-networks specifies the perceptions embedded by online users around the same entities, for example, coronavirus, in social discourses coming from different contexts.

3.1. Conceptual Relevance in Hashtag Networks

The largest connected components of the three hashtag networks included: 1000 hashtags and 8923 links for *#italylockdown*; 720 hashtags and 5915 links for *#sciacalli*; 6665 hashtags and 53395 links for *#italylockdown*. All three networks are found to be highly clustered (mean local clustering coefficient [39] of 0.82) and with an average distance between any two hashtags of 2.1. Only 126 hashtags were present in all the three networks.

Table 1 reports the most central hashtags included in each corpus of tweets thematically revolving around *#iorestoacasa*, *#sciacalli* and *#italylockdown*. Two rankings are considered: (i) hashtag frequency in tweets revolving around a key hashtag, and (ii) closeness centrality, which in here quantifies the tendency for hashtags to co-occur with other hashtags expressing analogous concepts and, therefore, are at short network distance from each other (see Section 2). Hashtags with a higher closeness centrality represent the prominent concepts in the social discourse. This result is similar to those showing that closeness centrality captures concepts which are relevant for early word acquisition [40] and production [47] in language. Additional evidence that closeness can capture semantically central concepts is represented by the closeness ranking, which assigns top-ranked positions to *#coronavirus* and *#COVID-19* in all three Twitter corpora. This is a consequence of the corpora being about the COVID-19 outbreak (and of the network metric being able to capture semantic relevance). Frequency and closeness rankings are only partially correlated, indicating that the two metrics highlight different hashtags of relevance in social discourse (Kendall τs: 0.213, $p < 10^{-6}$, for *#iorestoacasa*, 0.234, $p < 10^{-6}$, for *#sciacalli* and 0.225, $p < 10^{-6}$, for *#italylockdown*).

In the hashtag network built around *#italylockdown*, the most central hashtags are relative to the coronavirus, including a mix of negative hashtags such as *#pandemia* (English: "pandemic") and positive ones such as *#italystaystrong*. Frequency misses this combination and highlights only positive hashtags. Similarly, the hashtag network built around *#sciacalli* highlighted both positive (*#facciamorete* (English: "let's network") and negative (*#irresponsabili*—English: "irresponsible") hashtags. However, the social discourse around *#sciacalli* also featured prominent hashtags from politics, including references to specific Italian politicians, to the Italian Government, and hashtags expressing protest and shame towards the acts of a prominent Italian politician. Last but not least, closeness highlighted as prominent *#mascherine* or face masks, whereas frequency missed that hashtag. The social discourse around *#iorestoacasa* included many positive hashtags, eliciting hope for a better future and the need to act responsibly (e.g., *#andratuttobene* - English: "everything will be fine", or *#restiamoacasa* - English: "let's stay at home"). The most prominent hashtags in each network (cf. Table 1) indicate the prevalence of a positive social discourse around *#iostoacasa* and the percolation of strong political debate in relation to the negative topics conveyed by *#sciacalli*. However, we want to extend these punctual observations of negative/positive valences of single hashtags to the overall global networks. To achieve this, we use emotional profiling.

3.2. Emotional Profiling of Hashtag Networks

Hashtags can be composed of individual or multiple words. By extracting individual words from the hashtags of a given network, it is possible to reconstruct the emotional profile of the social

discourse around the focal hashtags *#sciacalli*, *#italylockdown* and *#iorestoacasa*. We tackle this by using the emotion-based [18] and the dimension based [31] emotional profiles (see Section 2).

Table 1. Top-ranked hashtags in co-occurrence networks based on closeness centrality (bottom) and word frequency (top). In all three rankings, the most central hashtag was the one defining the topic (e.g., *#italylockdown*) and was omitted from the ranking.

Frequency Rank	**#italylockdown**	**#sciacalli**	**#iorestoacasa**
1	iorestoacasa	Meloni	italylockdown
2	coronavirus	SanitaPubblica	coronavirus
3	COVID19	Sciacalli	italystaystrong
4	Covid19	Salvini	COVID19
5	IoRestoACasa	coronavirus	iostoacasa
6	andratuttobene	emergenzaCoronavirus	ItaliaZonaRossa
7	restiamoacasa	ItaliaViva	restiamoacasa
8	grazieanomeditutti	Sciacalle	CoronaVirusitaly
9	Coronavirus	governo	Coronavirus
10	coronavirusitalia	COVID19	Conte
11	andràtuttobene	Coronavirus	restaacasa
12	flashmob	terroristi	coranavirusit
13	covid19italia	irresponsabili	Covid19
14	litaliachiamò	Lega	celafaremo
15	Iorestoacasa	Berlusconi	Italy
16	coronarvirusitalia	facciamorete	COVID19italia
17	iostoacasa	FreeJulianAssange	italiazonaprotetta
18	coranavirusitalia	Conte	pandemia
19	covid19	CoronaVirus	coranavirusitalia
20	IORESTOACASA	coronavi	COVID2019
Closeness Rank	**#italylockdown**	**#sciacalli**	**#iorestoacasa**
1	coronavirus	coronavirus	iostoacasa
2	COVID19	COVID19	coronavirus
3	ItaliaZonaRossa	Salvini	COVID19
4	iorestoacasa	Lega	andratuttobene
5	Covid19	iorestoacasa	Covid19
6	Italia	Conte	restiamoacasa
7	italystaystrong	facciamorete	quarantena
8	Italy	Governo	COVID2019
9	pandemia	COVID2019	COVID19italia
10	COVID2019	Meloni	iorestoincasa
11	coronavirusitalia	Covid19	coronarvirusitalia
12	italiazonaprotetta	6marzo	italia
13	restiamoacasa	coronvirusitalia	andràtuttobene
14	lockdown	mascherine	covid19italia
15	iostoacasa	zonarossa	pandemia
16	coronarvirusitalia	Salvinivergognati	coranavirusitalia
17	coranavirusitalia	Coronavirus	CoronaVirusitaly
18	CoronaVirusitaly	coranavirusitalia	covid19
19	Conte	irresponsabili	Italia
20	COVID19italia	restiamoacasa	coronavirusitalIa

The emotional profiles of hashtags featured in co-occurrence networks are reported in Figure 2 (top). The top section of the figure represents perceived valence and arousal represented as a circumplex model of affect [31]. This 2D space or disk is called *emotional circumplex* and its coordinates represent emotional states that are well-supported by empirical behavioural data and brain research [31]. As explained also in the figure caption, each word is endowed with an (x, y) coordinate expressing its perceived valence (x) and arousal (y). Different points indicate different emotional combinations. For instance, (1,0) is the point of maximum/positive valence and zero arousal, that is, calmness; (0,−1) is the point of zero valence and minimum arousal, that is, lethargy; (−0.6,+0.6) represents a point of strong negative valence and positive arousal, that is, alarm.

Figure 2 reports the emotional profiles of all hashtags featured in co-occurrence networks for *#italylockdown* (left), *#sciacalli* (middle) and *#iorestoacasa* (right). To represent the interquartile range of all words for which valence/arousal rating are available, we use a neutrality range.

Histograms falling outside of the neutrality range indicate specific emotional states expressed by words included within hashtags (e.g., *#pandemia* contains the word "pandemia" with negative valence and high arousal).

3.2.1. Emotional Profiling of Hashtag Networks through the Circumplex Model

In Figure 2 (left, top), the peak of the emotional distribution for hashtags associated with *#italylockdown* falls within the neutrality range. This finding indicates that hashtags co-occurring with *#italylockdown*, a neutral hashtag by itself, were also mostly emotionally neutral conceptual entities. Despite this main trend, the distribution also features deviations from the peak mostly in the areas of calmness and tranquillity (positive valence, lower arousal) and excitement (positive valence, higher arousal). Weaker deviations (closer to the neutrality range) were present also in the area of anxiety.

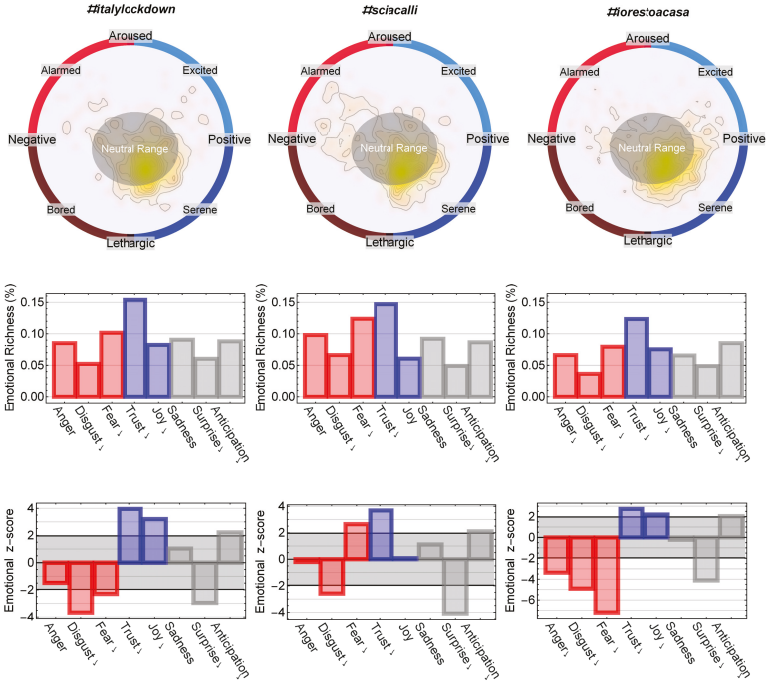


Figure 2. Emotional profiles of all hashtags featured in co-occurrence networks for *#italylockdown* (left), *#sciacalli* (middle) and *#iorestoacasa* (right). **Top:** Circumplex emotional profiling. All hashtags representing one or more words were considered. For each word, valence (x-coordinate) and arousal (y- coordinate) scores were attributed (Section 2) resulting in a 2D density histogram (yellow overlay) relative to the probability of finding an hashtag in a given location in the circumplex, the higher the probability the stronger the colour. Regions with the same probabilities are enclosed in grey lines. A neutrality range indicates where 50% of the words in the underlying valence/arousal dataset would fall and it thus serves as a reference value for detecting abnormal emotional profiles. Distributions falling outside of this range indicate deviations from the median behaviour (i.e., interquartile range, see Section 2). **Middle and Bottom:** NRC -based emotional profiling, detecting how many hashtags inspired a given emotion in a hashtag network. Emotions are colour-coded: red indicates negative emotions while blue and grey indicate positive and neutral emotions, respectively. Results are normalised over the total number of hashtags in a network. Emotions incompatible with random expectation ($|z - score| > 1.96$) were marked with a tick. Z-scores based on random samplings are reported at the bottom, together with the rejection region (grey overlay).

This reconstructed emotional profile indicates that the Italian social discourse featuring *#italylockdown* was mostly calm and quiet, perceiving the lockdown as a positive measure for countering responsibly the COVID-19 outbreak.

Not surprisingly, the social discourse around *#sciacalli* shows a less prominent positive emotional profile, with a higher probability of featuring hashtags eliciting anxiety, negative valence and increased states of arousal, as it can be seen in Figure 2 (center, top). This polarised emotional profile represents quantitative evidence for the coexistence of mildly positive and strongly negative content within the online discourse labelled by *#sciacalli*. This is further evidence that the negative hashtag *#sciacalli* was indeed used by Italian users to denounce or raise alarm over the negative implications of the lockdown, especially in relation to politics and politicians' actions. However, the polarisation of political content and debate over social media platforms has been encountered in many other studies [12,13,22] and cannot be attributed to the COVID-19 outbreak only.

Finally, Figure 2 (top right) shows that positive perception was more prominently reflected in the emotional profile of *#iorestoacasa*, which was the hashtag massively promoted by the Italian Government for supporting the introduction of the nationwide lockdown in Italy. The emotional profile of the 6000 hashtags co-occurring with *#iorestoacasa* indicate a considerably positive and calm perception of domestic confinement, seen as a positive tool to stay safe and healthy. The prominence of hopeful hashtags in association with *#iorestoacasa*, as reported in the previous subsection, indicate that many Italian Twitter users were serene and hopeful about staying at home at the start of lockdown.

3.2.2. Emotional Profiling of Hashtag Networks through Basic Emotions

Hashtag networks were emotionally profiled not only by using the circumplex model (see above) but also by using basic emotional associations taken from the NRC Emotion lexicon (Figure 2, bottom). Across all hashtag networks, we find a statistically significant peak in trust (z-scores > 1.96), analogous of the peaks close to emotions of calmness and serenity an observed in the circumplex models. However, hashtag networks included also negative emotions like fear, which is a natural human response to unknown threats and were observed also with the circumplex representations. All networks featured less disgust eliciting words than random expectation. The intensity of fearful, alarming and angry emotions is stronger in the *#sciacalli* hashtag network, which was used by social users to denounce, complain and express alertness about the consequences of the lockdown.

In addition to the politically-focused jargon highlighted by closeness centrality alone, by combining closeness with graph distance entropy (see Section 2 and Reference [33]) we identify other topics which are uniformly at short distance from others in the social discourse around *#sciacalli*, such as: *#mascherine* (English: "protective masks", which was also ranked high by using closeness only), *#amuchina* (the most popular brand, and synonym of, hand sanitiser), *#supermercati* (English: "supermarkets"). This result suggests an interesting interpretation of the negative emotions around *#sciacalli*. Beside the inflaming political debate and the fear of the health emergency, in fact, a *third* element emerges: Italian twitter users feared and were angry about the raiding and stockpiling of first aid items, symptoms of panic-buying in the wake of the lockdown.

3.3. Assessing Conceptual Relevance and Emotional Profiles of Hashtags via Word Networks

The above comparisons indicate consistency between dimension-based (i.e., the circumplex) and emotion-specific emotional profiling. Since the latter offers also a more precise categorisation of words in emotions, we will focus on emotion-specific profiling. Importantly, to fully understand the emotional profiles outlined above, it is necessary to identify the language expressed in tweets using a given combination of hashtags (see also Figure 1, bottom). As the next step of the MERCURIAL analysis, we gather all tweets featuring the focal hashtags *#italylockdown*, *#sciacalli*, or *#iorestoacasa* and any of their co-occurring hashtags and build the corresponding word networks, as explained in the Methods. Closeness centrality over these networks provided the relevance of each single word in

the social discourse around the topic identified by a hashtag. Only words with closeness higher than the median were reported.

Figure 3 shows the cloud of words appearing in all tweets that include *#sciacalli*, displayed according to their NRC emotional profile. Similar to the emotional profile extracted from hashtags co-occurring with *#sciacalli*, the words used in tweets with this hashtag also display a polarised emotional profile with high levels of fear and trust. Thanks to the multi-layer analysis, this dichotomy can now be better understood in terms of the individual concepts eliciting it.

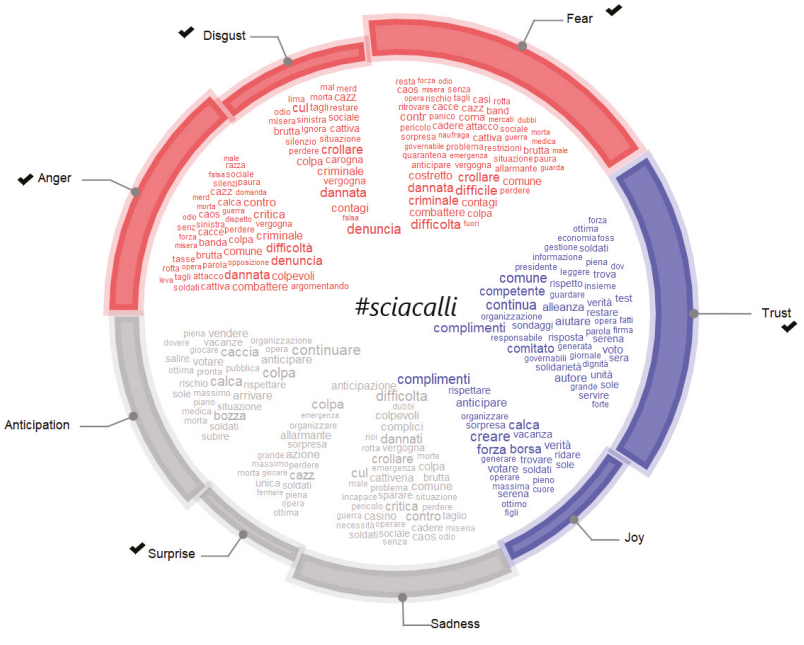


Figure 3. Emotional profile and word cloud of the language used in tweets with *#sciacalli*. Words are organised according to the emotion they evoke. Font size is larger for words of higher closeness centrality in the word co-occurrence network relative to the hashtag (see Section 2). Every emotion level incompatible with random expectation is highlighted with a check mark. Emotions are colour-coded: red indicates negative emotions, while blue and grey indicate positive and neutral emotions, respectively.

By using closeness on word networks, we identified concepts such as “competente” (English: “competent”), “continua” (English: “continue”, “keep going”), and ”comitato” (English: “committee”) to be relevant for the trust-sphere. These words convey trust in the expert committees appointed by the Italian Government to face the pandemic and protect the citizens. We find that other prominent words contributing to make the discourse around *#sciacalli* trustful are “aiutare” (English: “to help”), “serena” (English: “serene”), “rispetto” (English: “respect”) and ”verità” (English: “truth”), which further validate a trustful, open-minded and fair perception of the political and emergency debate outlined above. This perception was mixed with negative elements, mainly eliciting fear but also sadness and anger. The jargon of a political debate emerges in the word cloud of fear: “difficoltà‘ (English: “difficulty”), “criminale” (English: “criminal”), “dannati” (English: “scoundrel”), “crollare” (English: “to break down”), “banda” (English: “gang”), “panico” (English: “panic”) and ”caos” (English: “chaos”). These words indicate that Twitter users felt fear directed to specific targets. A speculative explanation for exorcising fear can be finding

a scapegoat and then target it with anger. The word cloud of such emotion supports the occurrence of such phenomenon by featuring words like "denuncia" (English: "denouncement"), "colpevoli" (English: "guilty"), "vergogna" (English: "shame"), "combattere" (English: "to fight") and "colpa" (English: "blame"). The above words are reflected also in other emotions like sadness, which features also words like "cadere" (English: "to fall") and "miseria" (English: "misery", "out of grace").

These prominent words in the polarised emotional profile of *#sciacalli*, suggest that Twitter users feared criminal behaviour, possibly related to unwise political debates or improper stockpiling of supplies (as showed by the hashtag analysis). Our findings also suggest that the reaction to such fearful state, which also projects sadness about negative economic repercussions, was split into a strong, angry denounce of criminal behaviour and messages of trust for the order promoted by competent organisations and committees. It is interesting to note that, according to Ekman's theory of basic emotions [24], a combination of sadness and fear can be symptomatic of desperation, which is a critical emotional state for people in the midst of a pandemic-induced lockdown.

The same analysis is reported in Figure 4 for the social discourse of *#italylockdown* (top) and *#iorestoacasa* (bottom). In agreement with the circumplex profiling, for both *#italylockdown* and *#iorestoacasa* the intensity of fear is considerably lower than trust.

However, when investigated in conjunction with words, the overall emotional profile of *#italylockdown* appears to be more positive, displaying higher trust and joy and lower sadness, than the emotional profile of *#iorestoacasa*. Although the difference is small, this suggests that hashtags alone are not enough to fully characterise the perception of a conceptual unit, and should always be analysed together with the natural language associated to them.

The trust around *#italylockdown* comes from concepts like "consigli" (English: "tips", "advice"), "compagna" (English: "companion", "partner"), "chiara" (English: "clear"), "abbracci" (English:"hugs") and "canta" (English: "sing"). These words and the positive emotions they elicit suggest that Italian users reacted to the early stages of the lockdown with a pervasive sense of commonality and companionship, reacting to the pandemic with externalisations of positive outlooks for the future, for example, by playing music on the balconies. This phenomenon was also mimicked in other countries later, and extensively reported by traditional media, see https://tinyurl.com/balconicovid, Last Access: 20 April 2020).

Interestingly, this positive perception co-existed with a more complex and nuanced one. Despite the overall positive reaction, in fact, the discourse on *#italylockdown* also shows fear for the difficult times facing the contagion ("contagi") and the lockdown restrictions ("restrizioni"), and also anger, identifying the current situation as a fierce battle ("battaglia") against the virus.

The analysis of anticipation, the emotional state projecting desires and beliefs into the future, shows the emergence of concepts such as "speranza" (English: "hope"), "possibile" (English: "possible") and "domani" (English: "tomorrow"), suggesting a hopeful attitude towards a better future.

The social discourse around *#iorestoacasa* brought to light a similar emotional profile, with a slightly higher fear towards being quarantined at home (quarantena (English: "quarantine"), comando (English: "command", "order", emergenza (English: "emergency"). Both surprise and sadness were elicited by the the word "confinamento" (English: "confinement"), which was prominently featured in the network structure arising from the tweets we analysed.

In summary, the above emotional profiles of hashtags and words from the 101,767 tweets suggest that Italians reacted to the lockdown measure with:

1. a fearful denunciation of criminal acts with political nuances and sadness/desperation about negative economic repercussions (from *#sciacalli*);
2. positive and trustful externalisations of fraternity and affect, combined with hopeful attitudes towards a better future (from *#italylockdown* and *#iorestoacasa*);
3. a mournful concern about the psychological weight of being confined at home, inspiring sadness and disgust towards the health emergency (from *#iorestoacasa*).

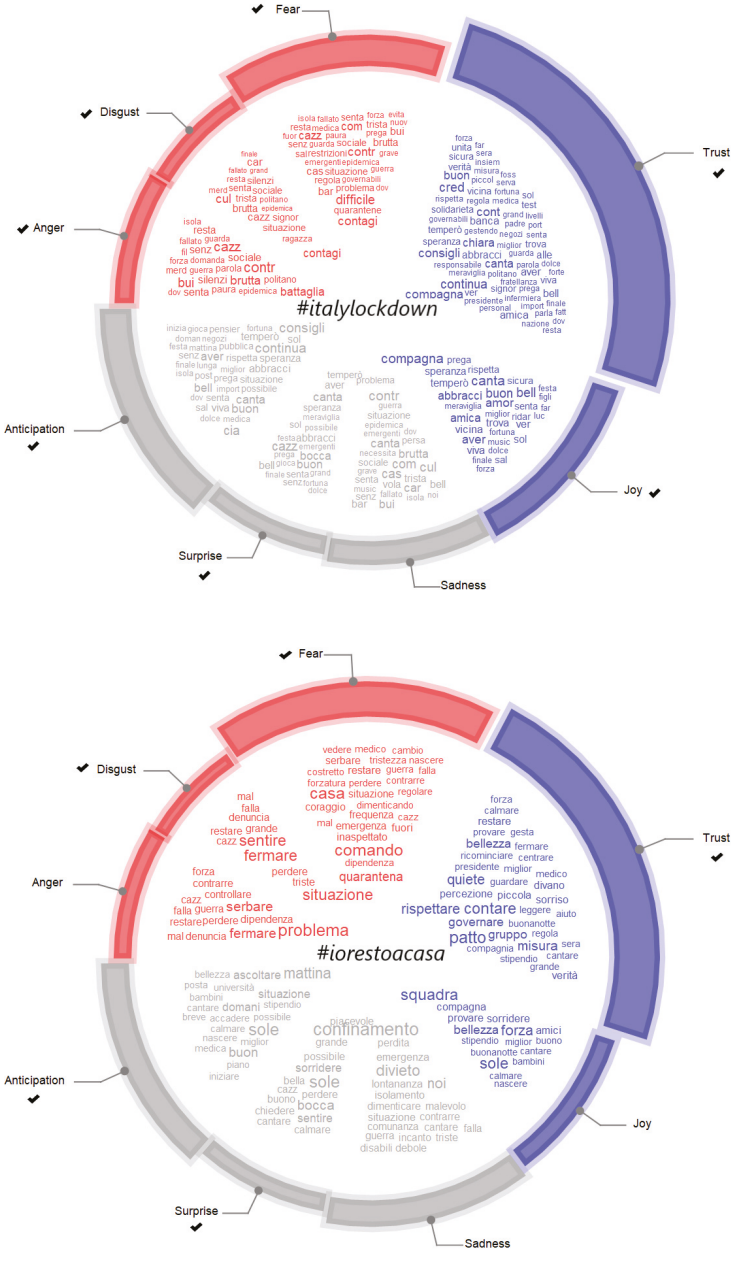


Figure 4. Emotional profile and word cloud of the language used in tweets with *#italylockdown* (**top**) and *#iorestoacasa* (**bottom**). Words are organised according to the emotion they evoke. Font size is larger for words of higher closeness centrality in the word co-occurrence network relative to the hashtag (Section 2). Every emotional richness incompatible with random expectation is highlighted with a check mark. Emotions are colour-coded: red indicates negative emotions, while blue and grey indicate positive and neutral emotions, respectively.

3.4. Hashtag Co-Occurrence Contextually Influences Hashtag Emotional Profiles

In the previous section we showed our findings on how Italians perceived the early days of lockdown on social media. But what about their perception of the ultimate cause of such lockdown, COVID-19? To better reconstruct the perception of *#coronavirus*, it is necessary to consider the different contexts where this hashtag occurs. Figure 5 displays the reconstruction of the emotional profile of words used in tweets with *#coronavirus* and either *#italylockdown*, *#sciacalli*, or *#iorestoacasa*.

Figure 5. Left: Emotional profiles with the NRC lexicon of the words occurring in tweets with the hashtag *#coronavirus* when it co-occurs with *#italylockdown* (outer circle), *#sciacalli* (middle circle) and *#iorestoacasa* (inner circle). **Right:** The same emotional profiles detected through the valence/arousal emotional circumplex. Every emotional richness incompatible with random expectation is highlighted with a check mark.

Our results suggest that the emotional profiles of language used in these three categories of tweets are different. For example, when considering tweets including *#sciacalli*, which the previous analysis revealed being influenced by political and social denounces of criminal acts, *#coronavirus* is perceived with a more polarised fear/trust dichotomy.

Although *#coronavirus* was perceived as trustful as random expectations when co-occurring with *#sciacalli* (z-score: 1.69 < 1.96), it was perceived with significantly higher trust when appearing in tweets with *#iorestoacasa* (z-score: 3.05 > 1.96) and *#italylockdown* (z-score: 3.51 > 1.96). To reinforce this picture, the intensity of fear towards *#coronavirus* was statistically significantly lower than

random expectation in the discourse of *#iorestoacasa* (z-score: $-2.35 < -1.96$) and *#italylockdown* (z-score: $-3.01 < -1.96$).

This difference is prominently reflected in both the circumplex model (Figure 5, right) and the NRC emotional profile (Figure 5, left), although in the latter both emotional intensities are compatible with random expectation. These quantitative comparisons provide data-driven evidence that Twitter users perceived the same conceptual entity, that is, COVID-19, with a higher trust when associating it to concrete means for hampering pathogen diffusion like lockdown and house confinement, and with a higher fear when denouncing the politics and economics behind the pandemic.

However, social distancing, lockdown and house confinement clearly do not have only positive sides. Rather, as suggested by our analysis, they bear complex emotional profiles, where sadness, anger and fear towards the current situation and future developments have been prominently expressed by Italians on social media.

4. Discussion

This study delved into the massive information flow of Italian social media users in reaction to the declaration of the pandemic status of COVID-19 by WHO, and the announcement of the nationwide lockdown by the Italian Government in the first half of March 2020. We explored the emotional profiles of Italians during this period by analysing the social discourse around the official lockdown hashtag promoted by the Italian Government (*#iorestoacasa*), together with a most trending hashtag of social protest (*#sciacalli*), and a general hashtag about the lockdown (*#italylockdown*).

The fundamental premise of this work is that social media opens a window on the minds of millions of people [17]. Monitoring social discourse on online platforms provides unprecedented opportunities for understanding how different categories of people react to real world events [9,12,15].

4.1. Impact of Cognitive Network Science for Social Media Analysis

Here we introduced a new framework, Multi-layer Co-occurrence Networks for Emotional Profiling (MERCURIAL), which is based on cognitive network science and that allowed us to:

(i) quantitatively structure social discourse as a multi-layer network of hashtag-hashtag and word-word co-occurrences in tweets;

(ii) identify prominent discourse topics through network metrics backed up by cognitive interpretation [14];

(iii) reconstruct and cross-validate the emotional profile attributed to each hashtag or topic of conversation through the emotion lexicon and the circumplex model of affect from social psychology and cognitive neuroscience [31].

Our interdisciplinary framework provides a first step in combining network and cognitive science principles to quantify sentiment for specific topics. Our analysis also included extensive robustness checks (e.g., selecting words based on different centrality measures, statistical testing for emotions), further highlighting the potential of the framework.

The analysis of concept network centrality identified hashtags of political denounce and protest against irrational panic buying (e.g., face masks and hand sanitiser) around *#sciacalli* but not in the hashtag networks for *#italylockdown* and *#iorestoacasa*. Our results also suggest that the social discourse around *#sciacalli* was further characterised by fear, anger, and trust, whose emotional intensity was significantly stronger than random expectation. We also found that the most prominent concepts eliciting these emotions revolve around social denounce (anger), concern for the collective well-being (fear), and the measures implemented by expert committees and authorities (hope). This interpretation is supported also by Plutchik's wheel of emotions [23], according to which combinations of anger, disgust and anticipation can be symptoms of aggressiveness and contempt. However, within Plutchik's wheel, trust and fear are not in direct opposition.

4.2. Evidence for Emotional Polarisation around COVID-19

The polarisation of positive/negative emotions observed around *#sciacalli* might be a direct consequence of a polarisation of different social users with heterogeneous beliefs, which is a phenomenon present in many social systems [22] but is also strongly present in social media through the creation of echo chambers enforcing specific narratives and discouraging the discussion of opposing views [2,10,11,13,48].

Emotional polarisation might therefore be a symptom of a severe lack of social consensus across Italian users in the early stages of the lockdown induced by COVID-19. In social psychology, social consensus is a self-built perception that the beliefs, feelings, and actions of others are analogous to one's own [49]. Destabilising this perception can have detrimental effects such as reducing social commitment towards public good or even lead to a distorted perception of society, favouring self-distrust and even conditions such as social anxiety [49]. Instead, acts such as singing from the balconies together can reduce fear and enhance self-trust [43], as well as promote commitment and social bonding [50], which is also an evolutionary response to help coping with a threat, in this case a pandemic, through social consensus. When interpreted under the lens of social psychology, the flash mobs documented by traditional media and identified here as relevant by semantic network analysis for *#italylockdown* and *#iorestoacasa* become important means of facing the distress induced by confinement [43,49,50].

4.3. Implications of the Detected Anger and Fear over Self-Awareness and Violence

Anger and fear permeated not only *#sciacalli* but were found, to a lesser extent, also in association with other hashtags such as *#iorestoacasa* or *#italylockdown*. Recent studies (cf. Reference [51]) found that anger and fear can drastically reduce individuals' sense of agency, a subjective experience of being in control of our own actions, linking this behavioural/emotional pattern also to alteration in brain states. In turn, a reduced sense of agency can lead to losing control, potentially committing violent, irrational acts [51]. Consequently, the strong signals of anger and fear detected here represent red flags about a building tension manifested by social users which might contribute to the outbreak of violent acts or end up in serious psychological distress due to lowered self-control.

One of the most direct implications of the detected strong signals of fear, anger and sadness is represented by increased violent behaviour. In cognitive psychology, the General Aggression Model (GAM) [52] is a well-studied model for predicting and understanding violent behaviour as the outcome of a variety of factors, including personality, situational context and the personal internal state of emotion and knowledge. According to GAM, feeling emotions of anger in a situation of confinement can strongly promote violent behaviour. In Italy, the emotions of anger and anxiety we detected through social media are well reflected in the dramatic rise in reported cases of domestic violence. For instance, the anti-violence centers of D.i.Re (*Donne in Rete Contro la Violenza*) reported an anomalous increase of +74.5% in the number of women looking for help for domestic violence in March 2020 in Italy (see the official report in Italian https://tinyurl.com/direcontrolaviolenza, Last Access: 20 April 2020). Hence, monitoring social media can be insightful about potential tensions mediated and discussed by large populations, a topic in need for further research and with practical prominent repercussions for fighting COVID-19.

4.4. Contextual Shifts in Emotions around the Novel Coronavirus

As discussed, we found the hashtag *#coronavirus* to be central across all considered hashtag networks. However, our analysis outlined different emotional nuances of *#coronavirus* across different networks. In psycholinguistics, contextual valence shifting [53] is a well-known phenomenon whereby the very same conceptual unit can be perceived wildly differently by people according to its context. This phenomenon suggests the importance of considering words in a contextual manner, by comparison to each other, as it was performed in this study, rather than alone. Indeed, contexts can

change the meaning and emotional perception of many words in language. We showed here that the same connotation shifting phenomenon [53] can happen also for hashtags. Online users perceived *#coronavirus* with stronger intensities of trust and lower fear (than random expectation) when using that hashtag in the context of *#iorestoacasa* and *#italylockdown*, but not when associated to *#sciacalli*. This shifting underlines the importance of considering contextual information surrounding a hashtag in order to better interpret its nuanced perception. To this aim, cognitive networks represent a powerful tool, providing quantitative metrics (such as graph distance entropy) that would be otherwise not applicable with mainstream frequency approaches in psycholinguistics.

4.5. Limitations and Future Research

MERCURIAL facilitates a quantitative characterisation of the emotions attributed to hashtags and discourses. Thanks to a better contextualisation of words by using hashtag-word multi-layer networks, MERCURIAL can also be used for all those tasks that are normally undertaken in NLP and revolve around gauging emotions from textual data. Some applications of MERCURIAL could include (but are not limited to) characterising emotional reactions to specific events, anticipate potential unrest or violence, or monitor the political discourse before an election.

Nonetheless, it is important to bear in mind that the analysis we conducted relies on some assumptions and limitations. For instance, following previous work [12], we built unweighted and undirected networks, neglecting information on how many times hashtags co-occurred. Including these weights would be important for detecting communities of hashtags, beyond network centrality. Notice that including weights would come at the cost of not being able to use graph distance entropy, which is defined over unweighted networks and was successfully used here for exposing the denounce of panic buying in *#sciacalli*. Another limitation is relative to the emotional profiling performed with the NRC lexicon, in which the same word can elicit multiple emotions. Since we measured emotional intensity by counting words eliciting a given emotion (plus the negations, see Section 2), a consequence was the repetition of the same words across the sectors of the above word clouds. Building or exploiting additional data about the predominance of a word in a given emotion would enable us to identify words which are peripheral to a given emotion, reduce repetitions and offer even more detailed emotional profiles. Recently, forma mentis networks [30,33] have been introduced as a method to detect the organisation of positive/negative words in the mindsets of different individuals. A similar approach might be followed for emotions in future research. Acting upon specific emotions rather than using the circumplex model would also solve another problem, in that the attribution of arousal to individual words is prone to more noise, even in mega-studies, compared to detecting word valence [54]. Notice also that the current analysis used uniform random sampling in the NRC lexicon for obtaining z-scores. This reference model keeps into account that some emotions might be elicited by fewer/more words than others but it also neglects potential idiosyncratic biases in the Italian twitter discourse. These biases should be detected and cross-validated through massive amounts of Italian twitter messages, beyond the dataset gathered here. The public release of massive multi-language datasets including Italian and reporting the COVID-19 Twitter discourse would enable more refined null models controlling also for idiosyncratic word under- or super-representation.

Another limitation is that emotional profiles might fluctuate over time. The insightful results outlined and discussed here were aggregated over a short time window, thus reducing the impact of aggregation itself. Future analyses on longer time windows should adopt time-series for investigating emotional patterns, addressing key issues like non-stationary tweeting patterns over time and statistical scarcity due to tweet crawling (see also Reference [12]). The current analysis has focused on aggregated tweets, but previous studies have shown both stable individual and intercultural differences in affect [55], especially for dimensions such as arousal. Similarly, some emotions are harder to measure than others, which might affect reliability and thus underestimate their contribution. The current approach estimates emotional profiles on the basis of a large set of words, which will reduce some

language-specific differences. The collection of currently missing large-scale Italian normative datasets for lexical sentiment could further improve the accuracy of the findings.

This study approaches the relation between emotions and mental distress mostly from the perspective that attitudes and emotions of the author are conveyed in the linguistic content. However, the emotion profile might also have implications for readers as well, as recent research suggests that even just reading words of strong valence/arousal can have deep somatic and visceral effects, for example, raising heart beat or promoting involuntary muscle tension [56]. Furthermore, authors and readers participate in an information network, and quantifying which tweets are liked or retweeted depending on the structure of social network can provide further insight on their potential impact [4,10,12,22,57], which calls for future approaches merging social networks, cognitive networks and emotional profiling.

Finally, understanding the impact of nuanced emotional appraisals would also benefit from investigating how these are related to behavioural and societal outcomes including the numbers of the contagion (e.g., hospitalisations, death rate, etc.) and compliance with physical distancing [58].

5. Conclusions

Given the massive attention devoted to the COVID-19 pandemic by social media, monitoring online discourse can offer an insightful thermometer of how individuals discussed and perceived the pandemic and the subsequent lockdown. Our MERCURIAL framework offered quantitative readings of the emotional profiles among Italian twitter users during early COVID-19 diffusion. The detected emotional signals of political and social denounce, the trust in local authorities, the fear and anger towards the health and economic repercussions, and the positive initiatives of fraternity, all outline a rich picture of emotional reactions from Italians. Importantly, the psychological interpretation of MERCURIAL's results identified early signals of mental health distress and antisocial behaviour, both linked to violence and relevant for explaining increments in domestic abuse. Future research will further explore and consolidate the behavioural implications of online cognitive and emotional profiles, relying on the promising significance of our current results. Our cognitive network science approach offers decision-makers the prospect of being able to successfully detect global issues and design timely, data-informed policies. Especially under a crisis, when time constraints and pressure prevent even the richest and most organised governments from fully understanding the implications of their choices, an ethical and accurate monitoring of online discourses and emotional profiles constitutes an incredibly powerful support for facing global threats.

6. Data Availability

The IDs of the tweets analysed in this study are available on the Open Science Foundation repository: https://osf.io/jy5kz/.

Author Contributions: Conceptualization, M.S. and V.R.; methodology, M.S. and S.D.D.; validation, M.S., V.R. and S.D.D.; formal analysis, M.S.; investigation, M.S., V.R. and S.D.D.; resources, M.S.; data curation, M.S.; writing—original draft preparation, M.S., V.R. and S.D.D.; writing—review and editing, M.S., V.R. and S.D.D.; visualization, M.S.; supervision, M.S.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: M.S. acknowledges Daniele Quercia, Nicola Perra and Andrea Baronchelli for stimulating discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zarocostas, J. How to fight an infodemic. *Lancet* **2020**, *395*, 676. [CrossRef]
2. Cinelli, M.; Quattrociocchi, W.; Galeazzi, A.; Valensise, C.M.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The covid-19 social media infodemic. *arXiv* **2020**, arXiv:2003.05004.
3. Gallotti, R.; Valle, F.; Castaldo, N.; Sacco, P.; De Domenico, M. Assessing the risks of "infodemics" in response to COVID-19 epidemics. *arXiv* **2020**, arXiv:2004.03997.
4. Pulido, C.M.; Villarejo-Carballido, B.; Redondo-Sama, G.; Gomez, A. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *Int. Soc.* **2020**. [CrossRef]
5. Wang, C.; Pan, R.; Wan, X.; Tan, Y.; Xu, L.; Ho, C.; Ho, R. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1729. [CrossRef]
6. World Health Organization. Mental Health During COVID-19 Outbreak. *Lancet Psychiatry* **2020**, *7*, e15–e16. [CrossRef]
7. Zhu, S. Wu, Y.; Zhu, C.; Hong, W.; Yu, Z.; Chen, Z.; Wang, Y. The immediate mental health impacts of the COVID-19 pandemic among people with or without quarantine managements. *Brain Behav. Immun.* **2020**, in press. [CrossRef]
8. Wang, C.; Pan, R.; Wan, X.; Tan, Y.; Xu, L.; Mclntyre, R.; Choo, F.; Tran, B.; Ho, R.; Sharma, V.; et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain Behav. Immun.* **2020**, in press. [CrossRef]
9. Ferrara, E.; Yang, Z. Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Comput. Sci.* **2015**, *1*, e26. [CrossRef]
10. Davis, J.T.; Perra, N.; Zhang, Q.; Moreno, Y.; Vespignani, A. Phase transitions in information spreading on structured populations. *Nat. Phys.* **2020**, *16*, 590–596. [CrossRef]
11. Ciulla, F.; Mocanu, D.; Baronchelli, A.; Gonçalves, B.; Perra, N.; Vespignani, A. Beating the news using social media: The case study of American Idol. *EPJ Data Sci.* **2012**, *1*, 8. [CrossRef]
12. Stella, M.; Ferrara, E.; De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12435–12440. [CrossRef]
13. Bail, C.A.; Argyle, L.P.; Brown, T.W.; Bumpus, J.P.; Chen, H.; Hunzaker, M.F.; Lee, J.; Mann, M.; Merhout, F.; Volfovsky, A. Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9216–9221. [CrossRef] [PubMed]
14. Siew, C.S.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive Network Science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, *2019*, 2108423. [CrossRef]
15. Stella, M. Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media. *arXiv* **2020**, arXiv:2003.08835.
16. Amancio, D.R. Probing the topological properties of complex networks modeling short written texts. *PLoS ONE* **2015**, *10*, e0118394. [CrossRef]
17. Quercia, D.; Kosinski, M.; Stillwell, D.; Crowcroft, J. Our twitter profiles, our selves: Predicting personality with twitter. In *Proceedings of the 2011 IEEE 3rd International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*; IEEE: Piscataway, NJ, USA, 2011; pp. 180–185.
18. Mohammad, S.M.; Turney, P.D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 26–34.
19. Dodds, P.S.; Harris, K.D.; Kloumann, I.M.; Bliss, C.A.; Danforth, C.M. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE* **2011**, *6*, e26752. [CrossRef]
20. Mohammad, S.; Bravo-Marquez, F.; Salameh, M.; Kiritchenko, S. Semeval-2018 task 1: Affect in tweets. In Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, LA, USA, 5–6 June 2018; pp. 1–17.
21. Kleinberg, B.; van der Vegt, I.; Mozes, M. Measuring Emotions in the COVID-19 Real World Worry Dataset. *arXiv* **2020**, arXiv:2004.04225.

22. Brito, A.C.M.; Silva, F.N.; Amancio, D.R. A complex network approach to political analysis: Application to the Brazilian Chamber of Deputies. *PLoS ONE* **2020**, *15*, e0229928. [CrossRef]
23. Plutchik, R. *The Emotions*; University Press of America: Lanham, MA, USA, 1991.
24. Ekman, P.E.; Davidson, R.J. *The Nature of Emotion: Fundamental Questions*; Oxford University Press: Oxford, UK, 1994.
25. Hatfield, E.; Cacioppo, J.T.; Rapson, R.L. Emotional contagion. *Curr. Direct. Psychol. Sci.* **1993**, *2*, 96–100. [CrossRef]
26. Barsade, S.G. The ripple effect: Emotional contagion and its influence on group behavior. *Adm. Sci. Quart.* **2002**, *47*, 644–675. [CrossRef]
27. Kramer, A.D.; Guillory, J.E.; Hancock, J.T. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8788–8790. [CrossRef] [PubMed]
28. Frey, S.; Donnay, K.; Helbing, D.; Sumner, R.W.; Bos, M.W. The rippling dynamics of valenced messages in naturalistic youth chat. *Behavi. Res. Methods* **2019**, *51*, 1737–1753. [CrossRef] [PubMed]
29. Jasper, J.M. Emotions and social movements: Twenty years of theory and research. *Ann. Rew. Soc.* **2011**, *37*, 285–303. [CrossRef]
30. Stella, M.; De Nigris, S.; Aloric, A.; Siew, C.S. Forma mentis networks quantify crucial differences in STEM perception between students and experts. *PLoS ONE* **2019**, *14*, e0222870. [CrossRef]
31. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. [CrossRef] [PubMed]
32. De Deyne, S.; Kenett, Y.N.; Anaki, D.; Faust, M.; Navarro, D.J. Large-scale network representations of semantics in the mental lexicon. In *Big Data in Cognitive Science: From Methods to Insights*; Routledge/Taylor & Francis Group: Abingdon, UK, 2016; pp. 174–202.
33. Stella, M.; Zaytseva, A. Forma mentis networks map how nursing and engineering students enhance their mindsets about innovation and health during professional growth. *PeerJ Comput. Sci.* **2020**, *6*, e255. [CrossRef]
34. Mehler, A.; Gleim, R.; Gaitsch, R.; Hemati, W.; Uslu, T. From Topic Networks to Distributed Cognitive Maps: Zipfian Topic Universes in the Area of Volunteered Geographic Information. *Complexity* **2020**, *2020*, 4607025. [CrossRef]
35. Stella, M.; De Domenico, M. Distance entropy cartography characterises centrality in complex networks. *Entropy* **2018**, *20*, 268. [CrossRef]
36. Chen, E.; Lerman, K.; Ferrara, E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Pub. Health Surv.* **2020**, *6*, e19273. [CrossRef]
37. Marinho, V.Q.; Hirst, G.; Amancio, D.R. Labelled network subgraphs reveal stylistic subtleties in written texts. *J. Complex Netw.* **2018**, *6*, 620–638. [CrossRef]
38. Vankrunkelsven, H.; Verheyen, S.; Storms, G.; De Deyne, S. Predicting lexical norms: A comparison between a word association model and text-based word co-occurrence models. *J. Cogn.* **2018**, *1*, 45. [CrossRef] [PubMed]
39. Newman, M. *Networks*; Oxford University Press: Oxford, UK, 2018.
40. Stella, M. Modelling early word acquisition through multiplex lexical networks and machine learning. *Big Data Cogn. Comput.* **2019**, *3*, 10. [CrossRef]
41. Warriner, A.B.; Kuperman, V.; Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **2013**, *45*, 1191–1207. [CrossRef] [PubMed]
42. Russell, J.A. A circumplex model of affect. *J. Pers. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
43. Unwin, M.M.; Kenny, D.T.; Davis, P.J. The effects of group singing on mood. *Psychol. Music* **2002**, *30*, 175–185. [CrossRef]
44. Fonagy, P.; Gergely, G.; Jurist, E.L. *Affect Regulation, Mentalization and the Development of the Self*; Routledge: Abingdon, UK, 2018.
45. Stella, M. Forma mentis networks reconstruct how Italian high schoolers and international STEM experts perceive teachers, students, scientists, and school. *Educ. Sci.* **2020**, *10*, 17. [CrossRef]
46. Miller, G.A. *WordNet: An Electronic Lexical Database*; MIT Press: Cambridge, MA, USA, 1998.
47. Castro, N.; Stella, M. The multiplex structure of the mental lexicon influences picture naming in people with aphasia. *J. Compl. Netw.* **2019**, *7*, 913–931. [CrossRef]

48. Brugnoli, E.; Cinelli, M.; Quattrociocchi, W.; Scala, A. Recursive patterns in online echo chambers. *Sci. Rep.* **2019**, *9*, 20118. [CrossRef]
49. Krueger, J. On the perception of social consensus. In *Advances in Experimental Social Psychology*; Elsevier: Amsterdam, The Netherlands, 1998; Volume 30, pp. 163–240.
50. Pearce, E.; Launay, J.; Dunbar, R.I. The ice-breaker effect: Singing mediates fast social bonding. *R. Soc. Open Sci.* **2015**, *2*, 150221. [CrossRef]
51. Christensen, J.F.; Di Costa, S.; Beck, B.; Haggard, P. I just lost it! Fear and anger reduce the sense of agency: A study using intentional binding. *Exp. Brain Res.* **2019**, *237*, 1205–1212. [CrossRef] [PubMed]
52. DeWall, C.N.; Anderson, C.A.; Bushman, B.J. The general aggression model: Theoretical extensions to violence. *Psychol. Viol.* **2011**, *1*, 245. [CrossRef]
53. Polanyi, L.; Zaenen, A. Contextual valence shifters. In *Computing Attitude and Affect in Text: Theory and Applications*; Springer: Berlin, Germany, 2006.
54. Mohammad, S.M. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL), Melbourne, VIC, Australia, 15–20 July 2018.
55. Kuppens, P.; Tuerlinckx, F.; Yik, M.; Koval, P.; Coosemans, J.; Zeng, K.J.; Russell, J.A. The relation between valence and arousal in subjective experience varies with personality and culture. *J. Pers.* **2017**, *85*, 530–542. [CrossRef] [PubMed]
56. Vergallito, A.; Petilli, M.A.; Cattaneo, L.; Marelli, M. Somatic and visceral effects of word valence, arousal and concreteness in a continuum lexical space. *Sci. Rep.* **2019**, *9*, 20254. [CrossRef]
57. Thelwall, M.; Thelwall, S. Retweeting for COVID-19: Consensus building, information sharing, dissent, and lockdown life. *arXiv* **2020**, arXiv:2004.02793.
58. Bedford, J.; Enria, D.; Giesecke, J.; Heymann, D.L.; Ihekweazu, C.; Kobinger, G.; Lane, H.C.; Memish, Z.; Oh, M.D.; Schuchat, A.; et al. COVID-19: Towards controlling of a pandemic. *Lancet* **2020**, *395*, 1015–1018. [CrossRef]

big data and cognitive computing

MDPI

Article

A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19

Toni Pano and Rasha Kashef *

Electrical, Computer, and Biomedical Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada; toni.pano@ryerson.ca

* Correspondence: rkashef@ryerson.ca

Received: 2 October 2020; Accepted: 30 October 2020; Published: 9 November 2020

Abstract: During the COVID-19 pandemic, many research studies have been conducted to examine the impact of the outbreak on the financial sector, especially on cryptocurrencies. Social media, such as Twitter, plays a significant role as a meaningful indicator in forecasting the Bitcoin (BTC) prices. However, there is a research gap in determining the optimal preprocessing strategy in BTC tweets to develop an accurate machine learning prediction model for bitcoin prices. This paper develops different text preprocessing strategies for correlating the sentiment scores of Twitter text with Bitcoin prices during the COVID-19 pandemic. We explore the effect of different preprocessing functions, features, and time lengths of data on the correlation results. Out of 13 strategies, we discover that splitting sentences, removing Twitter-specific tags, or their combination generally improve the correlation of sentiment scores and volume polarity scores with Bitcoin prices. The prices only correlate well with sentiment scores over shorter timespans. Selecting the optimum preprocessing strategy would prompt machine learning prediction models to achieve better accuracy as compared to the actual prices.

Keywords: sentiment analysis; Twitter; COVID-19; VADER scoring; correlation

1. Introduction

Recent research studies have emerged that involve the impact of COVID-19 on the financial market, including cryptocurrencies [1–8]. It was illustrated that Bitcoin is not a safe haven [1,2]. A correlation between Bitcoin and the stock market is observed in [3,4]. In [5], authors performed a dynamic correlation analysis that illustrated that Bitcoin could not hedge the US stocks' extraordinary tail risk. The co-movement between Bitcoin and daily data of COVID-19 world deaths is examined in [6]. The herding behavior in the cryptocurrency market has been explored in [7]. The association between the stock market volatility and policy responses to the COVID-19 outbreak is studied in [8]. Previous work before the pandemic has used various machine learning algorithms to predict the bitcoin price. In [9], Arti Jain et al. attempted to predict the prices of Bitcoin and Litecoin two hours in advance based on the sentiments expressed in current tweets. They investigated if social factors could predict the prices of cryptocurrencies. They used a Multiple Linear Regression (MLR) model to predict a bihourly average price from the number of positive, neutral, and negative tweets accumulated every two hours. Symeonidis et al. compared the significance of different preprocessing techniques for sentiment analysis of tweets [10]. They used four different machine learning algorithms, Linear Regression, Bernoulli Naïve Bayes, Linear Support Vector Machine, and a Convolutional Neural Network to classify tweets as positive, negative, or neutral sentiment. They tested 16 different preprocessing techniques in isolation. They recommended using lemmatization, replacing repeated punctuation, replacing contractions, or removing numbers. They identified the five most impactful techniques for use in a second. Based on their results, they suggested replacing URLs and user mentions, replacing contractions, replacing repeated punctuation, and lemmatization for a neural network classification

model. Ibrahim et al. [11] provided a predictive model to the BTC prices using Bayesian autoregression models. In Tan and Kashef [12], a comparative study between various BTC prediction models is performed, showing the Multi-Layer Perceptron (MLP) efficiency in forecasting the Bitcoin price. None of the above research studies have examined the impact of the emotions expressed about bitcoin on social media platforms, such as Twitter, during the COVID-19 outbreak. The preprocessing of tweets is a significant challenge in providing and building an acute prediction model. Feeding text data that do not correlate well with Bitcoin to a prediction model will not allow the right forecasting of Bitcoin's behavior. The purpose of this paper is to perform a complete Valence Aware Dictionary and sEntiment Reasoner (VADER)-based sentiment analysis of BTC tweets during the era of COVID-19 to identify the role of different preprocessing strategies in predicting Bitcoin prices. The sentiment analysis includes converting tweet text into a sentiment score that is representative of its emotion. Such a task is suited to VADER, a lexicon and rule-based sentiment analysis tool that can deal with the syntax usually adopted on social media. We developed 13 different preprocessing strategies for BTC tweets. To rank the preprocessing strategy's effectiveness, an overall correlation value, the Average Feature Correlation Magnitude (AFCM), is constructed. For each strategy, the correlation values of all optimum features are averaged by their absolute value. The optimum preprocessing strategies are quantified using VADER scoring systems. The VADER score is used to match the actual BTC price trend. Among all strategies, it was found that splitting sentences, removing Twitter-specific tags, or their combination generally improve the correlation of sentiment scores and volume polarity scores with Bitcoin prices. The rest of this paper is organized as follows: In Section 2, a discussion on sentiment analysis is presented; Section 3 presents related work on tweets sentiments. In Section 4, a Complete Sentiment Analysis using VADER scoring of BTC Tweets during the era of COVID-19 is discussed; Section 5 concludes the paper and provides future research directions.

2. Sentiment Analysis

In this section, some of the well-known sentiment analysis methods are discussed, including VADER [13], Word2vec [14], TFIDF [15], and N-grams [16].

2.1. Valence Aware Dictionary and sEntiment Reasoner (VADER)

VADER is a lexicon- and rule-based sentiment analysis tool that can handle words, abbreviations, slang, emoticons, and emojis commonly found in social media [13]. It is typically much faster than machine learning algorithms, as it requires no training [13,17]. Each body of text produces a vector of sentiment scores with negative, neutral, positive, and compound polarities [13]. The negative, neutral, and positive polarities are normalized to be between 0 and 1. The compound polarity can be thought of as an aggregate measure of all the other sentiments, normalized to be between −1 (negative) and 1 (positive).

2.2. Word to Vector (Word2vec)

In the Word2vec method, embedding words are designed as vectors made of real-valued numbers [14]. These vectors preserve both the syntactic regularity by making similar words have similar vectors and the semantic regularity between word pairs through vector algebra. For example, the word vector for "man" subtracted from the word vector for "king" and added to the word vector for "woman" produces something very similar to the word vector for "queen". Such vectors provide a way of translating words to numbers for train machine learning algorithms while preserving the relationships between words. Word2vec was initially developed by Tomas Mikolov et al. [14] to produce high-quality word vectors more efficiently than conventional means. A shallow feed-forward neural network trains weights with only input, projection, and output layers as word vectors. Two different network architectures, CBOW and Skip-gram, were used to generate the word vectors. CBOW is optimized to guess a target word from adjacent words in the text, while Skip-gram is optimized to guess adjacent words surrounding a target word.

2.3. Term Frequency-Inverse Document Frequency (TF-IDF)

The TFIDF approach finds keywords for each document in a collection of documents [15,16,18]. It assigns a number to each word in a document based on how frequently it appears in that document and how many documents use it. The "term-frequency" of a word is the number of times that word appears in a document, while the "document frequency" is the number of documents that contain that word [18]. The "inverse document frequency" of a word is the natural logarithm of the total number of documents divided by the word's "document frequency". Each word is assigned a TFIDF score by multiplying the word's "term frequency" by its "inverse document frequency".

2.4. N-Gram

N-gram is a model describing the identification of all groups of n adjacent words in a body of text [16]. For example, all adjacent triplets' words in the sentence "Mary had a little lamb" are "Mary had a", "had a little", and "a little lamb". These groupings are known as trigrams. However, groupings can be defined for any integer size. Unigrams, or single word groups, would include "Mary", "had", "a", "little,", and "lamb". Bigrams, or pairs of words, would include "Mary had", "had a", "a little", and "little lamb". Using bigger n-grams in a sentiment dictionary may help improve the accuracy of sentiment analysis when handling negations.

3. Related Work and Background

In [19], a correlation between sentiment analysis using current Bitcoin tweets and future Bitcoin price fluctuations was investigated. Based solely on sentiment changes, the naive prediction model achieves 83% accuracy with very few predictions. Authors in [20] show that Twitter sentiment and message volume could predict the price fluctuations of multiple cryptocurrencies, while Twitter bot accounts could potentially spread cryptocurrency misinformation. A modified VADER algorithm classified the tweet sentiments of nine cryptocurrencies' as buying, holding, or selling. Their experimental outcome showed that the daily intervals of Twitter sentiments and message volumes are stronger predictors than the buying to selling ratio. T. R. Li et al. [21] have attempted to demonstrate that Twitter's sentiments help in predicting cryptocurrency price changes. They have trained an Extreme Gradient Boosting Regression tree model (XGBoost) with Twitter sentiments to predict price changes. Six hourly variables for positive, negative, neutral, unweighted, retweet weighted sentiments, and trading volume were produced from the collected datasets. In [22], The VADER sentiment analysis algorithm was used to assign each tweet a compound sentiment score based on how positive, negative, or neutral their words were. The final sentiment score factored in the number of Twitter followers, likes, and retweets associated with each tweet. The current closing price of Bitcoin, final sentiment score, and the moving average of the last 100 data points were used as input variables for the model. C. Kaplan et al. [23] researched if rumors and speculation in social media can influence cryptocurrencies and price changes. Precisely, they gauged the dependence between the unstable cryptocurrency prices on Twitter sentiments. The six cryptocurrencies chosen were Agrello, Bread, Bytecoin, Digibyte, Doge coin, and Icos. Regression analysis was performed to test the dependence of daily cryptocurrency prices on daily Twitter sentiment. Significance F and R^2 values were calculated for each cryptocurrency. Bread and Bytecoin showed the lowest R^2 scores, while other coins had scores above 0.22. Agrello, Bytecoin, and Icos all had prominent F scores below 0.05. They concluded that some unstable cryptocurrencies might show dependence on Twitter sentiments. Sailunaz and Alhajj [24] created user recommendations for Twitter Users or topics. They showed that analyzing the full text from tweets proved to be better than exploring full text from tweets with only nouns, adjectives, verbs, and adverbs (NAVA). Their work involved providing sentiment scores, a reply network, and a follower network from the tweets to estimate machine learning recommendations. A Naïve Bayes classifier proved to work better than a Support Vector Machine (SVM) or a Random Forest (RF) under k-fold cross-validation [25]. The sentiment scores of the full text were a minimum of 5% better than NAVA text under 3-, 5-, and 10-fold cross-validation. The best score

was 66.86%, obtained from the 10-fold cross-validation of a Naïve Bayes classifier on full text. Hanjia Lyu et al. [26] characterized Twitter users who use controversial terms when mentioning COVID-19 on Twitter and trained various machine learning algorithms for classifying users.

4. A Complete Sentiment Analysis of BTC Tweets During the Era of COVID-19

This paper aims to identify modifications on the tweet text during preprocessing so that the resulting sentiment scores best correlated with Bitcoin's closing prices. We created different ways of preprocessing text for VADER scoring and tested them on truncated and full-length tweets.

4.1. Data Collection

We gathered tweets for sentiment analysis by developing a custom tweet scraper using Twitter API. We chose to collect data for three main reasons manually. All existing online free datasets did not include the COVID-19 pandemic period. All web scrapers were avoided because they might bypass the restrictions of the Twitter API. These restrictions were meant to protect Twitter users. We followed twitters rules [27,28], and we coded our tweet scraper in Python using the Tweepy library to access the Twitter API [29]. In our experiments, the collection method obtained a representative set of BTC tweets during the COVID-19 period. The tweet selection involved filtering tweets by a manually chosen set of keywords. Tweets that contained any keywords related to bitcoin ("bitcoin", "bitcoins", "Bitcoin", "Bitcoins", "BTC", "XBT", and "satoshi") or any hashtags of Bitcoin's ticker symbols ("#XBT", "$XBT", "#BTC", and "$BTC") were collected. Raw tweet text and their timestamps were stored. Timestamps were provided at a temporal resolution to the nearest second. As Twitter truncates tweets over 140 characters, the full-length version of those tweets was also collected [30–32]. A total of 4,169,709 tweets were collected from 8:47 AM, 22 May to 11:59 PM, 10 July. The volume of tweets collected for each date was observed to vary based on how old the requested data are, as shown in Figure 1. Bitcoin prices are obtained for free from the CryptoCompare API [32]. They provide open historical data of opening, high, low, and closing prices and volume (OHLCV) information at a temporal resolution of every minute [33]. Minutely, Bitcoin data were obtained over hourly data to provide enough data points to analyze. About 71,472 min of data points was collected from 22 May to 10 July, while collecting hourly prices would have provided nearly 1191 data points. Timestamps of Bitcoin prices and OHLCV data were then stored. The recorded OHLCV data from (Cryptocompare.com) seemed to fluctuate when prices were still recent. Data were provided up to 33 h into the past (based on our tests). A bi-daily collection routine was used to replace any recent prices (near the start of the collection period) that matched timestamps with any older prices from the next collection period.

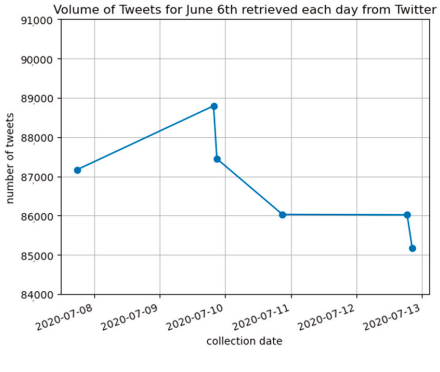


Figure 1. The volume of tweets created on 6 June, collected each day from Twitter's API.

4.2. Data Preprocessing

Preprocessing was performed on the text from each tweet converted into an average polarity score and tweet polarity volume per minute. This involves combining three main text cleaning functions labeled "cleaned," "split,", and "no sw". Respectively, they managed the removal of tweet-specific syntax, splitting text into sentences, and removing stopwords. The "cleaned" and "split" functions were tested in different orders, with and without the presence of the "no sw" function at the end. All three preprocessing functions affected the VADER sentiment analysis of text in different ways. Each had the potential to significantly help VADER capture a different aspect of sentiment from the text. The "cleaned" function removed unwanted characters and words used specifically on Twitter's platform, such as hyperlinks, numbers, and tweet specific syntax, using regular expressions. The removal was applied to preserve emojis and possible emoticon characters for use in the VADER sentiment analyzer. Before removing any alphanumeric chars, the ellipsis mark " … " was removed from the end of tweet text truncated to fit within 140 characters. Additionally, HTML entities such as "&" were converted to UTF-8 equivalent characters, such as "&". Then hyperlinks starting with the characters "http" or "www." were removed. Numbers, along with any symbols, punctuation, or units next to them, were removed. Finally, the tweet-specific syntax was removed. This syntax included mentions of usernames of the form "@username," hashtags of the form "#hashtag" and the start of retweets of the form "RT @username." Once the cleaning phase was completed as shown in Figure 2, each tweet was represented by words, whitespace, emojis, and other non-alphanumeric characters. Due to the difficulty of creating a regular expression to recognize all emoticons in VADER's lexicon [34], these characters were left unchanged. Therefore, the "cleaned" text attempted to leave everything that VADER could use in sentiment analysis unchanged.

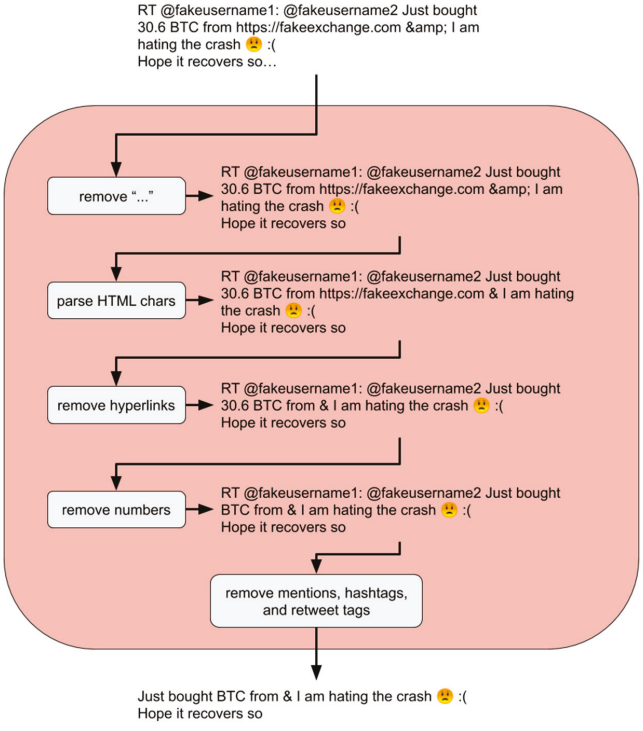


Figure 2. The preprocessing of the "cleaned" function over sample tweet text.

The "no sw" function in Figure 3 tokenizes text into words and removes any stopwords that VADER's dictionary does not use. Removing stop words from text requires a tokenized text into a list of words. The tokenization of text into words involves separating continuous blocks of alphabetical characters from the rest of the text. Blocks of continuous whitespace mark our word boundaries, split by Python's split () function [35]. Removing all non-alphabetical characters would solve this; however, this would remove some punctuation, all emojis, and all emoticons that VADER could recognize for sentiment analysis [36]. VADER allows exclamation marks "!" and question marks "?", which affect the sentiment score [36]. Our tokenization algorithm, as shown in Figure 4, groups characters from every tokenized word into "alphabetical", "punctuation", or "emoticon" blocks of characters. Ideally, these three blocks of characters would allow VADER to join each of them to preserve most of the text VADER can recognize. To distinguish punctuation from emoticons, any characters in the set $=@&_*#>:'\</{})]|%;~-,([+^" are only part of an emoticon if they occur next to another character in the same set.

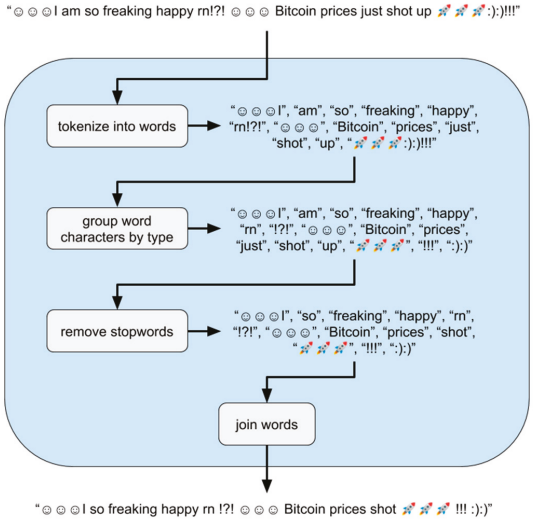


Figure 3. The preprocessing of the "no sw" function over sample text.

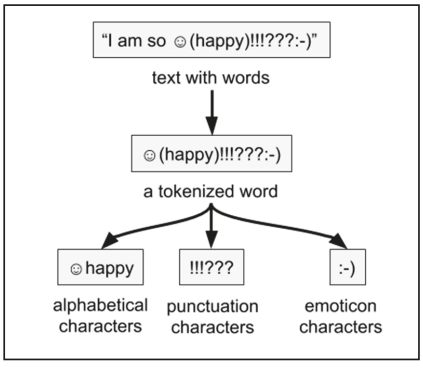


Figure 4. Tokenization of a word and character sorting.

To determine whether certain preprocessing functions contribute to predicting Bitcoin prices, the output of those functions in various combinations was scored by VADER. All text cleaning functions

of the preprocessing stage were combined in 5 different pathways. Scores of the text at intermediate steps of each path were recorded to determine if a function offers any improvement to the results. The "cleaned", "NLTK split", "regex split", and "no sw" functions shown in Figure 5 were combined to provide five different pathways. The "no sw" function was treated as the last step in any path, as no other function required word tokenization. Since VADER can be applied to a text of any length, the "cleaned", "NLTK split", and "regex split" functions produced a text of varying length. The "cleaned" and "split" functions were interchanged in different pathways. Both "split", "cleaned", and a "no sw" functions were applied afterward. The "cleaned" function can have either a "split" function used before the "no sw" stage. Our preprocessing combinations measured the scores of 13 intermediate steps (from 5 different pathways), as shown in Figure 6. The text tweets and the preprocessed dataset are available for access and download in [37].

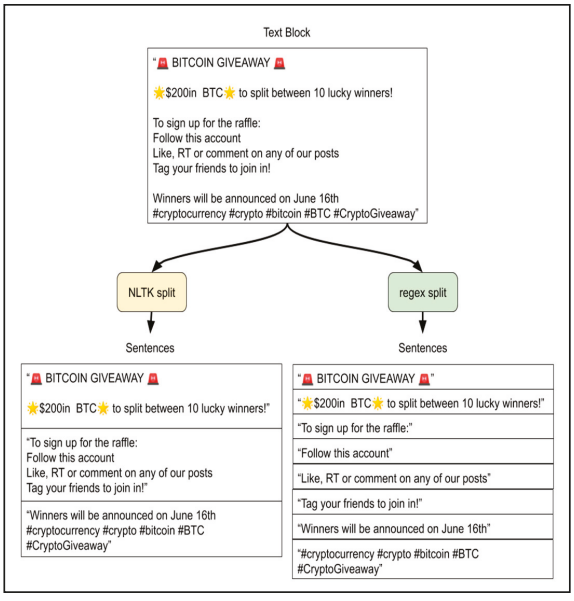


Figure 5. A comparison of sentence splitting by the regular expression and the NLTK sentence tokenizer.

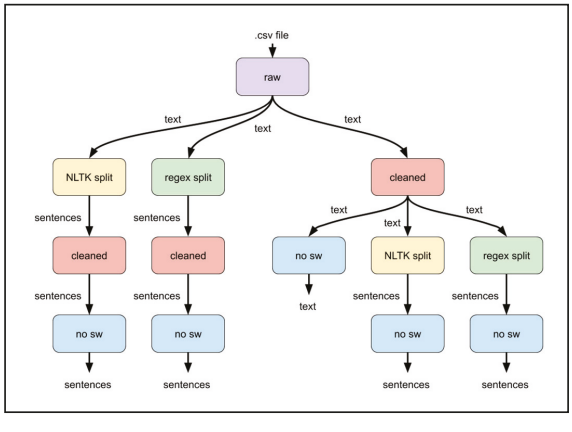


Figure 6. Five different preprocessing pathways with 13 collective intermediate steps (boxes).

4.3. VADER Sentiment and BTC Prices

This section uses the Pearson correlation between the VADER scores of each of the 13 intermediate preprocessing steps over time with BTC's closing prices per minute. As tweets are created much more frequently than once a minute, we aggregated all tweets' scores into a per-minute feature in two ways. First, we averaged the negative, neutral, positive, and compound scores of all tweets within each minute. This strategy allows us to preserve VADER's scoring property of the sentiment polarity scores summing to about 1. The second approach involves counting the number (or volume) of tweets that fit an overall sentiment polarity class. Each tweet can be classified as having an overall negative, neutral, or positive sentiment polarity based on its compound sentiment score. We consider any text with a compound VADER sentiment score below −0.05 as having an overall negative polarity, above 0.05, as having an overall positive polarity. Other scores have an overall neutral polarity. This produces 4 VADER sentiment score features and four sentiment volume features [13].

Since there were eight different features obtained for each of the 13 intermediate preprocessing steps over the whole dataset, the Bitcoin prices per minute were correlated against 104 unique time series of numbers. The resulting Pearson correlation score for each time series is displayed in a 13×8 heatmap matrix, as shown in Figure 7. This correlation matrix represents the Pearson correlation values of all types of time-series with Bitcoin prices. Each time-series was constructed from the full length of tweet text posted between 22 May, 8:47 AM, and 23 May, 8:47 AM. To rank the effectiveness of different preprocessing strategies, an overall correlation value, the Average Feature Correlation Magnitude (AFCM), was constructed for each matrix row. For each strategy, the correlation values of all eight features are averaged by their absolute value to produce a single value in the rightmost column. While a few patterns can be seen from the matrix, they can change if a shorter time length of data or a different time series start date is used. We improve the correlation matrix by graphing correlation over different dataset lengths in the next section.

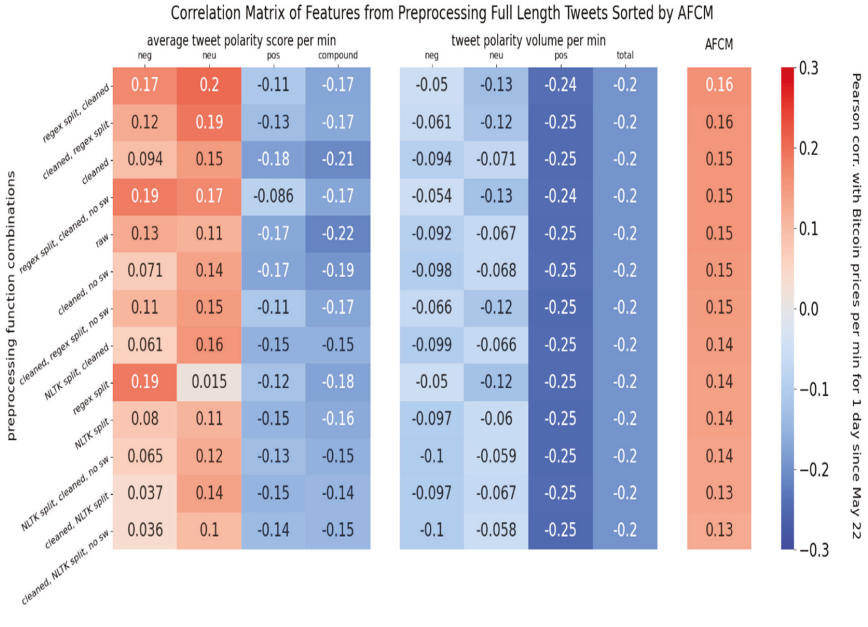


Figure 7. Correlation of sentiment timeseries and Bitcoin prices on 22 May.

4.4. Feature Types and Correlation

To account for differences in correlation due to the time length of the data used, we used subsets of data. The number of consecutive days of data varied and correlated with the respective Bitcoin prices occurring with the same timestamp. We will refer to this value as the correlation value of the subset. There are multiple unique subsets of data that span the same number of days. For example, a subset of 3 days of data can start on 22 May, 23 May, 24 May ... 5 July, 6 July, and 7 July. Therefore, we averaged the correlation values from all unique subsets of data with the same length and differing start dates. The resultant value is independent of its start date (as much as it can be with a finite time length of collected data). Averaging the correlation values of all same-length subsets with different start dates should show us any correlation polarity (positive or negative) that a majority of subsets show. This is known as the Average Subset Correlation Polarity (ASCP, dashed line in figures). The correlation value could be positive, negative, or averaging; thus, we might hide how large the correlation values were and make the ASCP approach zero. To mitigate this effect, we can also average the absolute correlation values of all same-length subsets with different start dates to show the magnitude or strength of the correlation values. This is known as the Average Subset Correlation Magnitude (ASCM, solid line in all figures). The ASCP and ASCM are plotted as line graphs against the length of data in all subsets. The following eight figures show the ASCP and ASCM for common features produced from all 13 preprocessing strategies. A performance ranking scheme graph is included in those figures to rank each strategy from best to worst using the ASCM, for all subset data lengths. The 1st rank corresponds to the best performance and largest ASCM, while the 13th rank corresponds to the worst performance and smallest ASCM. The Pearson correlation average might not be a good representation of correlation magnitude if both positive and negative correlation values are averaged. Therefore, another experimental analysis using the absolute correlation value showing the average correlation magnitude per timespan length was conducted. For a total of 49.6 days of data, we measured all possible subset outcomes, and then we spanned contiguous days of data. Any subsets with the same timespan length and differing start dates had their outcomes averaged. This produced a time-series of correlation values for each of the cells in our correlation matrix.

We graphed the correlation time-series for all preprocessing strategies that share the same aggregation score type in the same figure to display this data. Figure 8 shows the trends for the correlation of average negative VADER sentiment with Bitcoin prices for different data timespans. Most preprocessing strategies performed better than raw text when using less than 20 days of data and show a negative ASCP. The top-performing strategies involve cleaning and splitting sentences using the NLTK library in any order before removing their stopwords. A general pattern of combining text cleaning and sentence splitting in any order had a higher correlation than removing stopwords from those combinations. Splitting sentences without being combined with other functions performed worse than the latter two combinations. However, this trend was reversed when using more than 20 days of data, as a positive ASCP developed. Splitting sentences on their own performed better than removing stopwords from any order of cleaned sentence splitting, which served better than any order of cleaned sentence splitting. Few preprocessing strategies performed consistently better than raw text, such as splitting sentences using a regex when using 35 to 45 days of data. Thus, the correlation of average negative VADER sentiments per minute showed opposite trends for datasets of different time lengths. The effectiveness of using any preprocessing strategy over raw text decreased as more days of data were used. Cleaning and splitting text in any order on 20 days of data or less seemed to work best, while splitting raw text into sentences using regexes worked best on more extended datasets. The most significant dip in the ASCP of −0.105 occurred when correlating 8 to 15 days of data. The largest peak in the ASCP of 0.123 occurred when correlating about 40 to 49.6 days of data. Figure 9 shows the correlation of average neutral VADER sentiment with Bitcoin prices over different data timespans. The only preprocessing strategies that consistently performed better than using raw text were cleaning text, splitting text into sentences with the NLTK library, and splitting sentences using NLTK after cleaning text. In general, combining NLTK split sentences with cleaning in any order reduced its ASCM,

and removing stopwords from them reduced it further. Similarly, eliminating stopwords from cleaned text reduced its ASCM. The best performing preprocessing strategies for the average VADER neutral score per minute do not involve regex splitting or removing stopwords. The highest peak in both ASCP and ASCM for all strategies occurred when using about 6 to 13 days of data, excluding the rise for one day of data. This range showed a positive correlation of about 0.135. Figure 10 shows the correlation between average positive VADER sentiment with Bitcoin prices. Cleaning text outperformed when using 1 or 2 days of data. Splitting sentences by using a regex performed better when using more than 20 days of data. A general pattern of combining text splitting functions with cleaning reduced their ASCM, and removing stop words reduced them further. This indicates that the best performing preprocessing strategies for the average positive VADER sentiment per minute is raw text, followed by single functions (cleaning or sentence splitting on their own). The largest negative peaks in ASCP, of −0.12 and −0.10, occurred when using about 5 to 8 and 31 to 38 days of data, respectively.

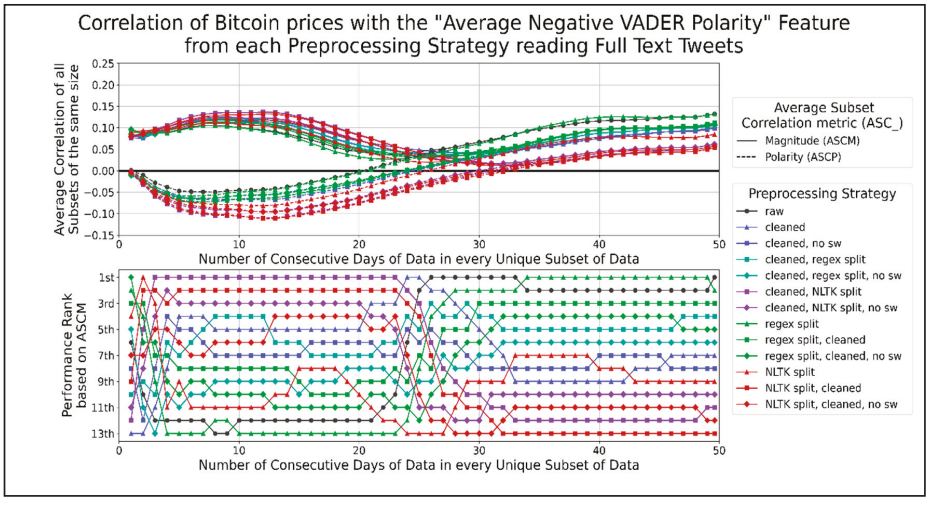


Figure 8. The correlation score time-series for the average negative VADER sentiment per minute.

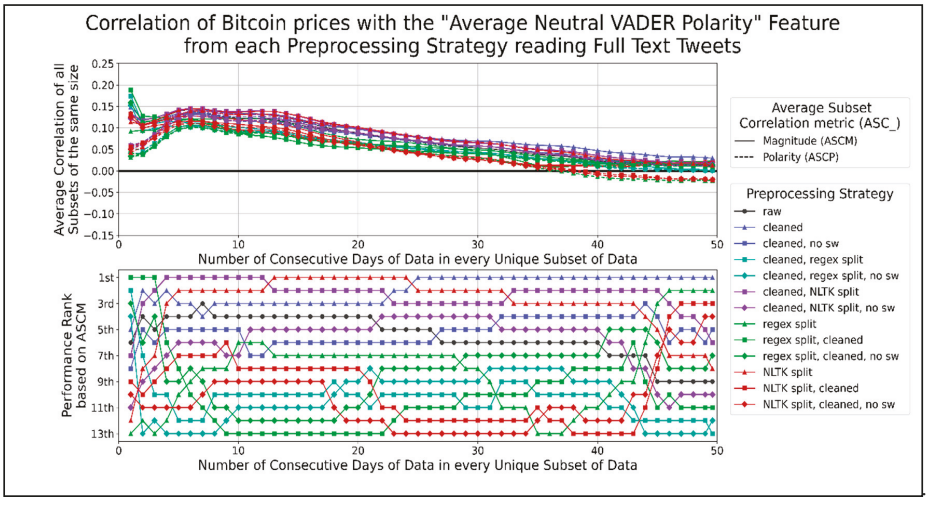


Figure 9. The correlation score time-series for the average neutral VADER sentiment per minute.

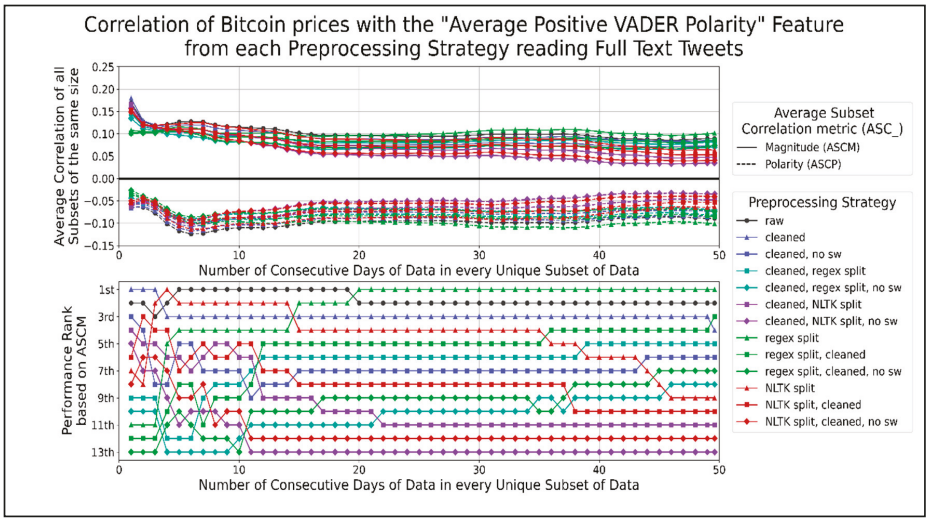


Figure 10. The correlation score time-series for the average positive VADER sentiment per minute.

Figure 11 shows correlation graphs of the average compound VADER sentiment at different timespans for the Bitcoin prices. No preprocessing strategies have a consistently higher ASCM than raw text. However, a few preprocessing strategies perform well for a few data subset lengths. Cleaned text, splitting sentences using the NLTK library, and splitting sentences using a regex performed better than using raw text when 1, 19, or more, and 34 or more data days were correlated. In Figure 11, in general, the best preprocessing strategies used the least amount of combined functions. Combinations of functions that split sentences using the NLTK library performed better than those using a regex. Therefore, the preprocessing strategies for the average compound VADER score per minute using raw text for less than 19 days of data and splitting sentences using the NLTK library for greater data lengths performed the best among other strategies. The largest negative peaks in ASCP were −0.06, −0.08, and −0.10 when correlating 1, 5 to 7, and 34 to 44 days of data, respectively.

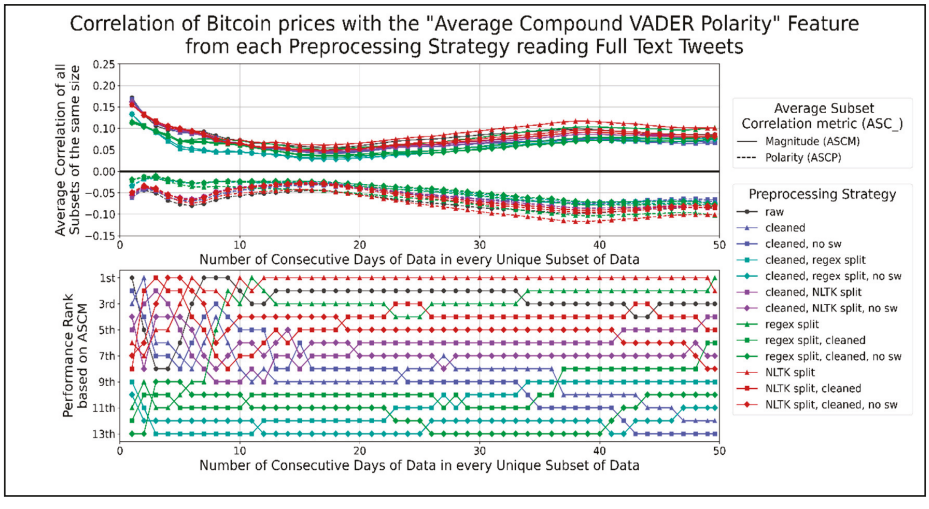


Figure 11. The correlation score time-series for the average compound VADER sentiment per minute.

Figure 12 shows the correlation graphs of negative tweets per minute with Bitcoin prices. The cleaned text and cleaned text with stopwords removed closely match the raw text correlation. Cleaned text performed better than raw text when using between 35 days and 47 days of data, coinciding with the largest peak in the ASCP. A general pattern of sentence splitting combined with text cleaning, with stopwords removed, performed better than any sole sentence splitting method, which performed better than any order of combining sentence splitting and cleaned text. There was a large gap in performance between the three top strategies: raw text, cleaned text, cleaned text without stopwords, and the other preprocessing strategies. Therefore, the ASCM of the volume of negative tweets per minute was the highest when using those three strategies. The highest peak in ASCP was 0.07, at 40 days of data. Another peak of about 0.055 occurred when using 4 to 6 days of data.

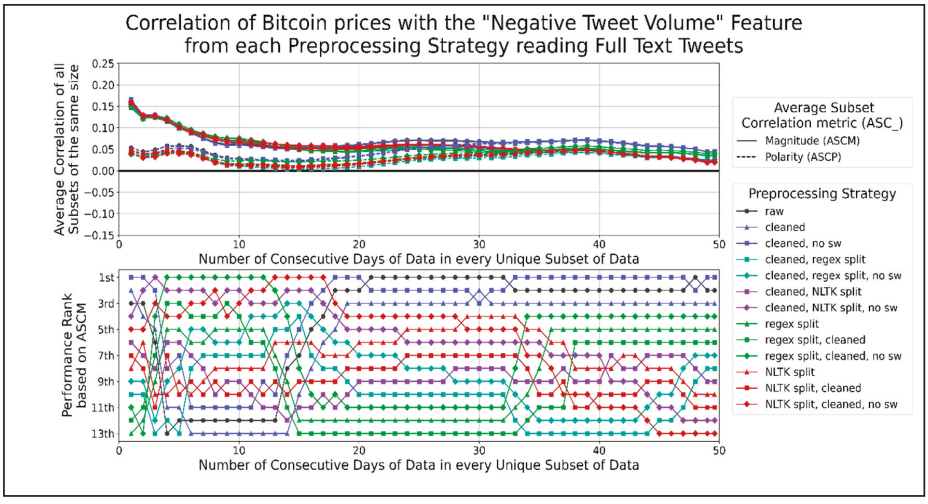


Figure 12. The correlation score time-series for the volume of negative tweets per minute.

Figure 13 shows the correlation graphs of neutral tweets per minute for Bitcoin prices at different timespans. The top preprocessing strategies are splitting sentences using the NLTK library or a regex. In general, strategies that do not use a regex to split sentences tend to follow raw text correlation closely. Preprocessing strategies that use less combined functions achieve a higher ASCM than otherwise. Therefore, the preprocessing strategy that allows the volume of neutral tweets per minute, which correlates the best with Bitcoin prices, is the NLTK library to split sentences. The highest peak in ASCP was 0.095 when using 6 to 13 days of data. Figure 14 shows the correlation of positive tweets per minute with Bitcoin prices. The preprocessing strategies that consistently performed better than using raw text were the sentence splitting functions combined with one or more functions, such as cleaning text and/or removing stopwords. The top two strategies that performed the best involve cleaning text before using a regex for sentence splitting. In general, preprocessing strategies that involve combining more functions perform better. Therefore, the best preprocessing strategy for correlating Bitcoin prices with the volume of positive tweets per minute involves cleaning text before sentence splitting by a regex function. The highest peak in ASCP was about 0.085 when using 12 to 20 days of data. Figure 15 shows the total correlation per minute of tweets for Bitcoin prices. No preprocessing strategy can affect the total amount of tweets received from Twitter per minute; hence every single preprocessing function would have ASCM and ASCP graphs, as in Figure 15. The highest peak in the ASCP was 0.09 and occurred when correlating 6 to 20 days of data.

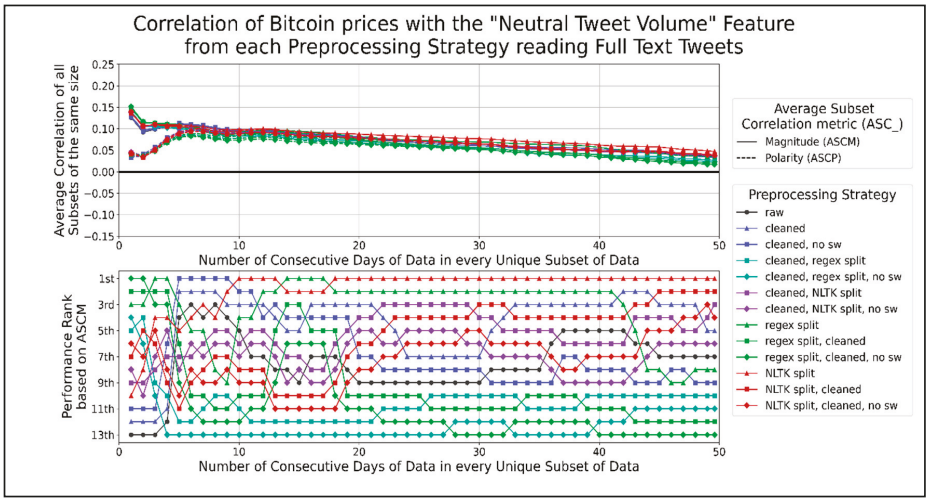


Figure 13. The correlation score time-series for the volume of neutral tweets per minute.

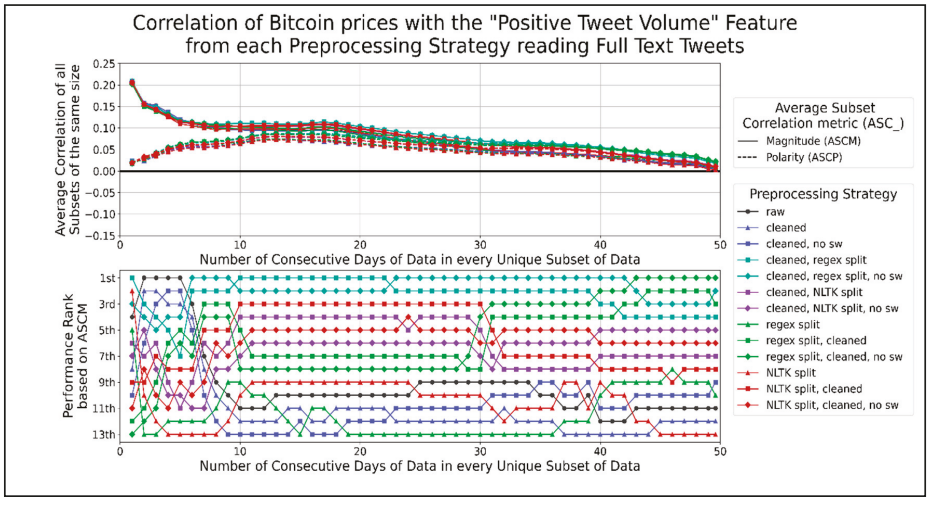


Figure 14. The correlation score time-series for the volume of positive tweets per minute.

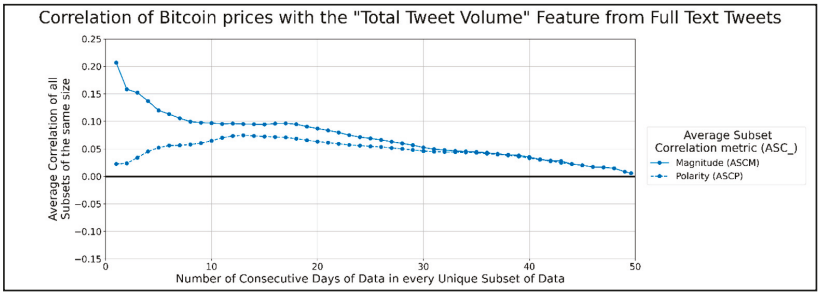


Figure 15. The correlation score time-series for the total amount of tweets posted per minute.

It is worth noting that the correlation graphs of the total volume of tweets have a similar trend as the correlation graphs of the neutral and positive volumes of tweets. This may indicate that the correlation of the total volume of tweets shares more in common with the volumes of the neutral and positive tweets than negative tweets when using long-term datasets on a scale of days. The general trend of the above graphs shows the strongest correlation magnitude for shorter datasets of full-text tweet data and Bitcoin prices. When using one day of data for correlation, all ASCMs are significantly higher than when using all other timespans of data (except the average negative VADER scores per minute graphs). This could be a sign of sentiments expressed on Twitter, either responding to or anticipating a Bitcoin price change. However, the ASCPs, when using one day of data, are significantly lower than the respective ASCMs. This might show that any correlation observed with sentiments varies a lot depending on the date. We speculate that the substantial spike in correlation magnitude for a day of data on every graph indicates that correlation may become even stronger when observed on a shorter timescale, such as minutes instead of days. While there is no single preprocessing strategy that performs better than the rest for all feature types, we can see that cleaning text (and/or) splitting sentences is presented in most of the best strategies of each features. Any sentence splitting by itself seemed to work best for average VADER positive/neutral sentiment and neutral tweet volume, while NLTK splitting combined with cleaning worked best for average VADER neutral sentiment. Any sentence splitting also worked best for neutral tweet volume, while cleaning text worked best for negative tweet volume, and regex sentence splitting after cleaning text worked best for positive tweet volume. The feature types with the highest ASCMs around their highest peak were the positive/total tweet volume when a day of data was processed. No clear "best" features could be seen when all lengths of data subsets were considered. The highest ASCMs were all average VADER sentiments when less than 20 days of data were in each subset. ASCPs tended to show a second peak when processing 35 or more days of data to calculate the average VADER negative/positive/compound sentiments and negative tweet volume. These peaks may indicate longer-term data trends that continue outside of our 49.6 days of data, but they are typically low. However, they may be useful in machine learning algorithms that account for the past state, as recurrent neural networks.

5. Discussion

In this paper, we contributed to BTC price forecasting literature by referring to the role of social media, namely Twitter messages, in the forecasting process [38]. It is most important to collect the tweets concerning people's thoughts, emotions, and opinions about BTC during the period. The resulting sentiment scores from various preprocessing strategies are used to calculate the correlation coefficient with actual Bitcoin prices during the era of COVID−19. Our results indicate that the strongest correlation comes from processing a day's worth of data, which has an unpredictable correlation polarity. However, some longer-term trends in correlation were observed when using ten days or 35+ data days, which might help machine learning algorithms that use temporal memory, such as recurrent neural networks. The patterns observed in processing full-length tweet text match closely with the patterns in processing truncated tweet text. This indicates that truncated tweet text is a suitable replacement for processing full-length tweet text at a reduced preprocessing cost, which helps process large datasets or real-time prediction systems such as KryptoOracle [22]. Although, in this paper, we used a short-term period for correlating tweets to BTC prices, the experimental period can be used as a sample for a long-term prediction for the entire period. We believe that we can apply the same preprocessing strategies to identify which one represents the BTC trend using the same set of strategies for the whole pandemic period. We expect that if the number of COVID-19 cases rises in specific regions and the tweets change, the proposed preprocessed strategies can work efficiently to provide well-fit representative data as long as we have the actual BTC price to compare with. Additionally, the collection period can be extended if the second wave of COVID-19 happens, such that a new corpus can be selected and matched to the actual BTC during the second wave. Recommendations for the optimum preprocessing strategies will be provided in the same trend provided in the experimental

work. Although computing correlations in short time windows provide a well-fit preprocessed model for better BTC forecasting, a significant challenge remains in the unpredictable correlation polarity in longer-term trends. Our preprocessing methods using correlations can successfully be used as a groundwork for knowledge modeling through cognitive networks such as neural networks and deep learning to work directly on the most representative preprocessed Twitter data to the actual BTC price for forecasting. Furthermore, the most-fit preprocessing model can be used to predict the BTC price trends effectively. This paper demonstrated the adoption of natal language processing to assess users and decision-makers in perceiving and monitoring Bitcoin [39].

6. Conclusions and Further Research

In this paper, we identified the optimal preprocessing strategy of Bitcoin tweets introduced in the VADER-based Sentiment Analysis during the era of COVID-19. This paper used the VADER score from text preprocessing strategies to relate to the Bitcoin prices trend in this era. In general, we observe that features from cleaning text of tweet syntax and splitting text into sentences, in combination or separately, somewhat correlate with Bitcoin prices. However, the best preprocessing strategy to use depends on the feature you wish to extract from the text. More complex strategies are not guaranteed to correlate better. It can be concluded that the VADER score from text preprocessing shows a significant short-term correlation with Bitcoin prices. Future research involves investigating how datasets with timespans at each minute, instead of each day, relate to this work. We would also like to investigate if the correlation of our features with Bitcoin prices, from any of our preprocessing strategies, indicates how well a machine learning algorithm performs for predicting the BTC price from those features. Further future directions would involve selecting and correlating the optimal preprocessing strategy after COVID-19, at which we expect that the emotions and opinions on Twitter would change.

Author Contributions: Conceptualization, T.P. and R.K.; methodology, T.P. and R.K.; software, T.P.; validation, R.K.; formal analysis, T.P. and R.K.; investigation, T.P. and R.K.; resources, T.P.; data curation, T.P. and R.K.; writing—original draft preparation, T.P. and R.K.; writing—review and editing, T.P. and R.K.; visualization, T.P.; supervision, R.K.; project administration, R.K.; funding acquisition, R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ryerson University, Faculty of Engineering Undergraduate Opportunity Fund and The APC was funded by Ryerson Start-up fund.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Conlon, T.; Richard, M. Safe haven or risky hazard? Bitcoin during the COVID-19 bear market. *Financ. Res. Lett.* **2020**, *35*, 101607. [CrossRef] [PubMed]
2. Kristoufek, L. Grandpa, Grandpa, Tell Me the One About Bitcoin Being a Safe Haven: New Evidence from the COVID-19 Pandemic. *Front. Phys.* **2020**, *8*. [CrossRef]
3. Corbet, S.; Charles, L.; Brian, L. The contagion effects of the COVID-19 pandemic: Evidence from gold and cryptocurrencies. *Financ. Res. Lett.* **2020**. [CrossRef]
4. Lahmiri, S.; Bekiros, S. The impact of COVID-19 pandemic upon stability and sequential irregularity of equity and cryptocurrency markets. *Chaos Solitons Fractals* **2020**, *138*, 109936. [CrossRef] [PubMed]
5. Grobys, K. When Bitcoin has the flu: On Bitcoin's performance to hedge equity risk in the early wake of the COVID-19 outbreak. *Appl. Econ. Lett.* **2020**, in press. [CrossRef]
6. Goodell, J.; Goutte, S. Co-movement of COVID-19 and Bitcoin: Evidence from wavelet coherence analysis. *Financ. Res. Lett.* **2020**. [CrossRef]
7. Yarovaya, L.; Matkovskyy, R.; Jalan, A. The Effects of a Black Swan Event (COVID-19) on Herding Behavior in Cryptocurrency Markets: Evidence from Cryptocurrency USD, EUR, JPY and KRW Markets. *SSRN Electron. J.* **2020**. [CrossRef]
8. Zaremba, A.; Kizys, R.; Aharon, D.Y.; Demir, E. Infected Markets: Novel Coronavirus, Government Interventions, and Stock Return Volatility around the Globe. *Financ. Res. Lett.* **2020**, *35*, 101597. [CrossRef]

9. Jain, A.; Tripathi, S.; Dwivedi, H.D.; Saxena, P. Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. In Proceedings of the 2018 Eleventh International Conference on Contemporary Computing (IC3) Institute of Electrical and Electronics Engineers (IEEE), Noida, India, 2–4 August 2018; pp. 1–7.
10. Symeonidis, S.; Effrosynidis, D.; Arampatzis, A. A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Syst. Appl.* **2018**, *110*, 298–310. [CrossRef]
11. Ibrahim, A.; Kashef, R.; Li, M.; Valencia, E.; Huang, E. Bitcoin Network Mechanics: Forecasting the BTC Closing Price Using Vector Auto-Regression Models Based on Endogenous and Exogenous Feature Variables. *J. Risk Fin. Manag.* **2020**, *13*, 189. [CrossRef]
12. Tan, X.; Kashef, R. Predicting the closing price of cryptocurrencies. In Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems-DATA '19, Association for Computing Machinery (ACM), Dubai, United Arab Emirates, 2–6 December 2019; pp. 1–5.
13. Hutto, C.J. VADER-Sentiment-Analysis, GitHub. Available online: https://github.com/cjhutto/vaderSentiment (accessed on 24 July 2020).
14. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781 [cs.CL].
15. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval. *J. Document.* **2004**, *60*, 493–502. [CrossRef]
16. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **2016**, *57*, 117–126. [CrossRef]
17. Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Available online: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109 (accessed on 24 July 2020).
18. Havrlant, L.; Kreinovich, V. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *Int. J. Gen. Syst.* **2017**, *46*, 27–36. [CrossRef]
19. Stenqvist, E.; Lönnö, J. Predicting Bitcoin Price Fluctuation with Twitter Sentiment Analysis. Bachelor's Thesis, School of Computer Science and Communication (CSC), KTH, Stockholm, Sweden, 2017.
20. Kraaijeveld, O.; De Smedt, J. The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *J. Int. Financial Mark. Inst. Money* **2020**, *65*, 101188. [CrossRef]
21. Li, T.R.; Chamrajnagar, A.S.; Fong, X.R.; Rizik, N.R.; Fu, F. Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model. *Front. Phys.* **2019**, *7*. [CrossRef]
22. Mohapatra, S.; Ahmed, N.; Alencar, P. KryptoOracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments. *arXiv* **2019**, arXiv:2003.04967 [cs.CL].
23. Kaplan, C.; Aslan, C.; Bulbul, A. Cryptocurrency Word-of-Mouth Analysis via Twitter, ResearchGate. 2018. Available online: https://www.researchgate.net/publication/327988035_Cryptocurrency_Word-of-Mouth_Analysis_viaTwitter (accessed on 20 May 2020).
24. Sailunaz, K.; Alhajj, R. Emotion and sentiment analysis from Twitter text. *J. Comput. Sci.* **2019**, *36*, 101003. [CrossRef]
25. Rosen, A. Tweeting Made Easier, Twitter. 7 November 2017. Available online: https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier.html. (accessed on 24 July 2020).
26. Lyu, H.; Chen, L.; Wang, Y.; Luo, J. Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19. *IEEE Trans. Big Data* **2020**, *1*. [CrossRef]
27. The Twitter Rules, Twitter. Available online: https://help.twitter.com/en/rules-and-policies/twitter-rules (accessed on 19 May 2020).
28. Automation rules, Twitter. Available online: https://help.twitter.com/en/rules-and-policies/twitter-automation (accessed on 19 May 2020).
29. Tweepy. 2009. Available online: http://www.tweepy.org/ (accessed on 19 May 2020).
30. Counting Characters, Twitter. Available online: https://developer.twitter.com/en/docs/basics/counting-characters (accessed on 24 July 2020).
31. Search Tweets-Overview-Search API, Twitter. Available online: https://developer.twitter.com/en/docs/tweets/search/overview/standard (accessed on 24 July 2020).
32. Search Tweets-API Reference-Standard search API, Twitter. Available online: https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets (accessed on 24 July 2020).

33. Choose Your Plan, CryptoCompare. Available online: https://min-api.cryptocompare.com/pricing (accessed on 24 July 2020).
34. Hutto, C.J. vaderSentiment/vaderSentiment/vader_lexicon.txt, GitHub. 22 May 2020. Available online: https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vader_lexicon.txt (accessed on 24 July 2020).
35. "5. Built-in Types"—Python 2.7.18 Documentation. Available online: https://docs.python.org/2/library/stdtypes.html (accessed on 24 July 2020).
36. Hutto, C.J. vaderSentiment/vaderSentiment/vaderSentiment.py, GitHub. 22 May 2020. Available online: https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vaderSentiment.py (accessed on 24 July 2020).
37. Pano, T.; Kashef, R. A Corpus of BTC Tweets in the Era of COVID-19. In Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Institute of Electrical and Electronics Engineers (IEEE), Vancouver, BC, Canada, 9–12 September 2020; pp. 1–4.
38. Stella, M. Modelling Early Word Acquisition through Multiplex Lexical Networks and Machine Learning. *Big Data Cogn. Comput.* **2019**, *3*, 10. [CrossRef]
39. Li, D.; Summers-Stay, D. Mapping Distributional Semantics to Property Norms with Deep Neural Networks. *Big Data Cogn. Comput.* **2019**, *3*, 30. [CrossRef]

big data and cognitive computing

MDPI

Article

Structural Differences of the Semantic Network in Adolescents with Intellectual Disability

Karin Nilsson [1,2,*], Lisa Palmqvist [1,2], Magnus Ivarsson [1,2], Anna Levén [1,2], Henrik Danielsson [1,2], Marie Annell [1,2], Daniel Schöld [1,2] and Michaela Socher [1,2]

[1] Department of Behavioural Sciences and Learning, Linköping University, 58183 Linköping, Sweden; lisa.palmqvist@liu.se (L.P.); magnus.ivarsson@liu.se (M.I.); anna.leven@liu.se (A.L.); henrik.danielsson@liu.se (H.D.); marie.e.annell@gmail.com (M.A.); daniel.schold@liu.se (D.S.); michaela.socher@liu.se (M.S.)

[2] Swedish Institute for Disability Research, Linköping University, 58183 Linköping, Sweden

* Correspondence: karin.a.nilsson@liu.se

Abstract: The semantic network structure is a core aspect of the mental lexicon and is, therefore, a key to understanding language development processes. This study investigated the structure of the semantic network of adolescents with intellectual disability (ID) and children with typical development (TD) using network analysis. The semantic networks of the participants (n_{ID} = 66; n_{TD} = 49) were estimated from the semantic verbal fluency task with the pathfinder method. The groups were matched on the number of produced words. The average shortest path length (ASPL), the clustering coefficient (CC), and the network's modularity (Q) of the two groups were compared. A significantly smaller ASPL and Q and a significantly higher CC were found for the adolescents with ID in comparison with the children with TD. Reasons for this might be differences in the language environment and differences in cognitive skills. The quality and quantity of the language input might differ for adolescents with ID due to differences in school curricula and because persons with ID tend to engage in different out-of-school activities compared to TD peers. Future studies should investigate the influence of different language environments on the language development of persons with ID.

Keywords: semantic network analysis; intellectual disability; adolescents

Citation: Nilsson, K.; Palmqvist, L.; Ivarsson, M.; Levén, A.; Danielsson, H.; Annell, M.; Schöld, D.; Socher, M. Structural Differences of the Semantic Network in Adolescents with Intellectual Disability. *Big Data Cogn. Comput.* **2021**, *5*, 25. https://doi.org/10.3390/bdcc5020025

Academic Editors: Massimo Stella and Yoed N. Kenett

Received: 23 April 2021
Accepted: 22 May 2021
Published: 1 June 2021

1. Introduction

The semantic network structure is a core aspect of the mental lexicon [1] and is, therefore, a key to understanding language development processes. Different methods have been applied to study the semantic network structure in various populations in recent years [2–5]. However, little is known about the semantic network structure in persons with intellectual disability (ID), although language limitations [6], including semantic verbal fluency deficits [7,8], are part of the ID symptomatology. A better understanding of the specific characteristics of the semantic networks in persons with ID can be an essential tool for the development of language interventions for the group. It may also give important clues about semantic network development in general by shedding light on the role of general intellectual functioning. The current study aimed to investigate if the semantic network structure in a sample of adolescents with ID and a control group of younger typically developing (TD) children, differs. Studying the structure of the semantic network may lead to important insight into the verbal profile of persons with ID. Differences in the structure could help to explain specific challenges seen in language ability [9] and memory [10] in the population with ID. Such knowledge could, in the long-term, lay the foundation for the development of more effective interventions aimed at strengthening different verbal abilities based on specific network features of the ID population.

Semantic network analysis builds on graph theory and offers new ways for analyzing how information such as words generated in verbal fluency tasks are stored in memory and later retrieved [11]. The basic elements of the semantic network are nodes (words) and edges (the relationships between the words). The edges represent the associative strength between words [12]. Words that are named in temporal proximity to each other are likely to be stored nearby in the mental space [13]. Data for the network analysis are often attained through a semantic fluency task, typically involving participants naming as many words as possible within a given category and time [14].

Different characteristics of semantic networks have been studied, including distances between nodes and the tendency and nature of the cluster formation. The shortest path length is defined as the minimum number of edges (steps) between two nodes. The average shortest path length (ASPL) is the average number of edges in the shortest path between all possible pairs of nodes [12]. A high ASPL indicates that the nodes are, on average, remotely connected in the semantic network. The clustering coefficient (CC) measures the extent to which the nodes and their neighboring nodes are interconnected [12,13]. A high CC indicates that the semantic network is densely clustered. Another common quantifier of a semantic network is the network's modularity (Q). Modularity is a measure of the tendency to form subgroups (communities) within the network [12,15]. A high Q indicates well-defined subgroups with many edges connecting nodes within the subgroups and few edges between nodes belonging to different subgroups [15]. Taken together, these three measures—ASPL, CC, and Q—describe the mental representation of the semantic network in an individual's long-term memory. See Figure 1 for visual representation of the three measures.

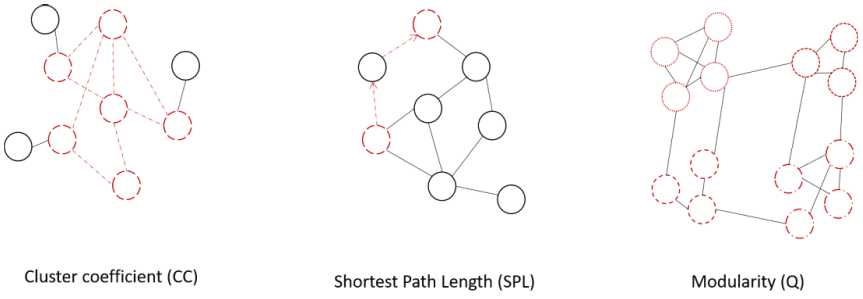


Figure 1. Visualization of the network measures used in the current study. The cluster coefficient measures the extent to which the nodes are interconnected. The shortest path length is the minimum number of edges between two nodes. The average shortest path length (ASPL) was calculated and used as a measure. Modularity measures the tendency to form subgroups within the network.

It has been argued that the structure of the semantic network might be the result of statistical learning, a process where taxonomic categories are formed based on co-occurrence regularities [16]. Evidence suggests that statistical learning is apparent in children as young as 4–5 years of age [16]. A prerequisite for statistical learning is that the similarities between contexts are detected and understood by the child, i.e., an ability to implicitly match patterns. Research studying statistical learning in persons with ID is sparse (see Saffran [17]), but it has been suggested that the capacity of implicit learning is functionally equivalent in young adults with and without ID [18] as well as in children and adolescents with and without ID matched on mental or chronological age [19]. However, Kover [20] argues that persons with ID may exhibit difficulties in implementing learning from distributional cues (i.e., patterns in input) and that weaker cognitive and linguistic skills may hinder efficient learning from cues. In addition, Thiessen et al. [21] suggest that the outcome of statistical learning changes during development as a function of experience and the maturity of the learner. Thus, it would be reasonable to assume that the ID

population differs from the TD population in terms of the outcome of statistical learning. If this is the case, and if statistical learning influences the structure of semantic networks, it follows that the semantic networks of persons with ID differ from those of TD peers. For example, semantic categories may be structured according to different principles in the ID and the TD groups. A specific word, such as "dog", could activate the category "house animals" in the TD group while activating random animals, words from other categories, or no other words at all in the ID group.

Statistical learning is likely not the only factor influencing the semantic network. In a conceptual framework for understanding the aging mental lexicon presented by Wulff et al. [1], learning processes are placed alongside aspects of the environment as factors that may affect the network structure. When it comes to environmental factors, Wulff et al. [1] suggest both qualitative (content) and quantitative (total amount of exposure) aspects that may be of importance. It might be the case that the environment differs between students with and without ID since the former group follows a different curriculum in Sweden [22] and tends to attend different out-of-school activities [23]. One aspect of learning highlighted by Wulff et al. [1] is that the encoding of new information is moderated by prior knowledge.

No previous study has applied network analysis to compare the semantic network structure of adolescents with ID to a TD sample based on data from a semantic fluency task. The current study will begin to fill this research gap. The number of words included in the network might influence the structure [24]. Therefore, controlling for size is essential. In the current study, this bias was reduced through the matching of groups based on the number of produced words on the semantic fluency task. The study aimed to answer the following research question: Does the structure of the semantic networks differ between adolescents with ID and children with TD, and if so, how?

Since prior research within the field of network structure in ID is scarce, there is no clear basis for formulating specific hypotheses, which motivates the explorative design of the present study. However, prior research and theories on statistical learning and semantic network development indicate the following interpretations of possible outcomes:

1. The chronological age of the ID group is higher compared with the TD group. Therefore, they should have been exposed to more language input, and their semantic network should be more developed than the semantic network of the comparison group, even if their total number of produced words are the same.
2. However, the limitations in cognitive functions might lead to the ID group not being able to make the same use of the language input as a TD group. Therefore, their semantic network might have a similar or less developed structure than the one of the comparison group.

2. Materials and Methods

The current study was an empirical study investigating differences in the semantic network between adolescents with ID and children with TD. Data from two different projects were used, and a network analysis using the pathfinder method was implemented. In the following sections, the sample, procedure, materials, and network analysis are described.

2.1. Participants and Recruitment

The verbal fluency data used in this paper are based on existing data from two different projects [25,26]. The data includes 49 participants with TD and 66 participants with ID. The participants with TD attended preschool class (i.e., Swedish school preparation class for children ~age 6). According to the teacher report, none of the children in the TD group had a developmental disability. The participants with ID were all adolescents, had an ID with unknown etiology, and attended compulsory school for students with ID. Participants in the ID group with additional disabilities, as indicated by a parental report, were excluded from the study.

Caregivers of all participants, and adolescents with ID over the age of 15, signed an informed consent form. Caregivers and participants were told they could drop out of the study without giving a reason. The data collection for the children with TD (Ref: 2015/308-31) and adolescents with ID (Ref: 2017/139-31) was approved by the local Research Ethics Review Committee in Linköping. The results from the semantic verbal fluency test of the TD group have been reported in another study about pragmatic language ability [27]. The result from the semantic verbal fluency data of the ID group is used in two pre-registered studies investigating reading ability in adolescents with ID [26].

2.2. Matching Procedure

The number of words included in the network might influence the structure [24]. Therefore, controlling for size is essential. This bias was reduced through the matching of groups based on the number of produced words on the semantic fluency task. Since the ID group was larger than the TD group, a subset of the ID group was selected for matching. One child from the TD group was excluded as the testing was disturbed several times. In addition, one outlier with a much higher score than the others from the TD group was excluded. Several, but not all, participants in the ID group could be individually matched to a child in the TD group. To minimize the effect of selection bias, ten different (but overlapping) samples were selected from 1,000,000 randomly generated possible selections from the ID group. These ten samples were chosen since they deviated the least from the mean (12.00) and standard deviation (4.08) of the number of produced animals in the TD group. All ten ID samples were matched on mean and SD to the TD group on the number of produced animals ($p > 0.65$; all means = 12.00, SD = 4.08 (range 4.07–4.09)). Both the TD group and the 10 different ID groups contained 47 participants each (64 different ID participants belonged to at least one selected ID group). The samples were comparable in terms of gender (21 females in the TD group; 22.5 females across the ID groups (range 20–25)). The TD group (M = 6:6 years, SD = 3.9 months) had a lower chronological age than the ID group (M = 15:11 years, SD = 27.6 months). All analyses were performed on all ten datasets. The results were pooled across datasets, and the pooled results are presented.

2.3. Semantic Verbal Fluency Test

The semantic network was assessed using the animal category subtest from the Delis-Kaplan Executive Function System (D-KEFS; [14]). The participants were asked to verbalize as many animals as possible in one minute. The total number of correctly generated words was used to match the groups. Fictional animals or duplicates were marked as invalid responses.

2.4. Procedure

The testing was conducted one-to-one in a quiet room at school (both groups) or at home (for some of the children in the TD group). The test administrator was a speech and language pathologist (both groups), an experienced test leader with a background in education (ID group), or a researcher with a background in cognitive science (TD group). All test sessions were recorded, and the recordings were used to transcribe the answers for the semantic verbal fluency task. The testing was part of two larger research projects. The children with TD as well as the adolescents with ID were tested on more tasks than are reported here.

2.5. Network Analysis

Only responses produced by at least two participants were included in the network analysis. We estimated the networks using the pathfinder method (see [13] for the recommendation, see [28,29] for method). The pathfinder method has been recommended for the estimation of group networks that are connected using every response, and networks are based on edge similarity [13]. The topological properties of the networks were validated using a bootstrapped random network analysis [24]. To compare the network of the ID

group and the TD group, a case-drop bootstrap analysis with 2000 runs was performed. This analysis was performed with all 10 datasets. ASPL, CC, and Q were thereafter pooled. Because the visualization of the networks cannot be pooled, the comparison was done for all 10 datasets with ID that gave the same results. To improve the readability of the results section, only one ID network is presented (for the dataset closest to the median values on a combination of ASPL, CC, and Q).

All analyses were performed in R [30] using the R packages tidyverse [31], readxl [32], tictoc [33], beepr [34], stringr [35], flextable [36], SemNetCleaner, SemNetDictionaries, and SemNeT [37]. An alpha level of 0.05 was used.

3. Results

3.1. Number of Unique Words

The adolescents with ID produced 129 (pooled result: 128.9) unique animals. Of these animals, 43.6 (39.8%) were only produced by one child. The children with TD produced 106 unique animals. Of these animals, 43 (40.6%) were produced by only one child.

3.2. Network Validation

ASPL, CC, and Q of all networks differed significantly from random ($p < 0.001$). The results of the random comparison are reported in Appendix A.

3.3. Network Comparison

The results from the pooled bootstrap analyses are reported in Table 1. A significantly smaller ASPL and Q were found for the adolescents with ID in comparison with the children with TD. In addition, a significantly higher CC was found for the adolescents with ID in comparison with the children with TD.

Table 1. The result from the pooled bootstrap analyses for ASPL, CC, and Q.

	ID Mean	TD Mean	$F(1,1997)$	p	Partial $\eta 2$	Direction
ASPL	2.73	2.84	178	<0.001	0.08	ID < TD
CC	0.48	0.43	800	<0.001	0.27	ID > TD
Q	0.35	0.37	308	<0.001	0.12	ID < TD

As can be seen in Figures 2 and 3, the network of the adolescents with ID includes more close nodes and exhibits a shorter ASPL. In addition, the network is less spread out. Neither the network of the children with TD nor the network of the adolescents with ID appear to have clear subgroups. Rather, many words are not clearly separated. For the network of the children with TD, there is a tendency towards the development of subgroups, which is not the case as much for the network of the adolescents with ID. This is also mirrored in the Q, which is significantly larger for the TD group compared with the ID group. Further, the adolescents with ID exhibit a higher CC compared with the children with TD. This is visible in the figures, as the largest subgroup of the ID group is more densely clustered (see bottom left in Figure 2) compared with the largest subgroup of the TD group (see top right in Figure 3). For the TD group, the developing subgroups are mostly related to the expected taxonomic structure of the animal category, while this is true to a lesser extent for the ID group. Note that the network layout was created using the Fruchterman–Reingold algorithm [37], which is very sensitive to small differences in network properties such as path length. The position of the large subgroups on the opposite ends for the ID group compared with the TD group is therefore merely an artifact of how the network plot was created.

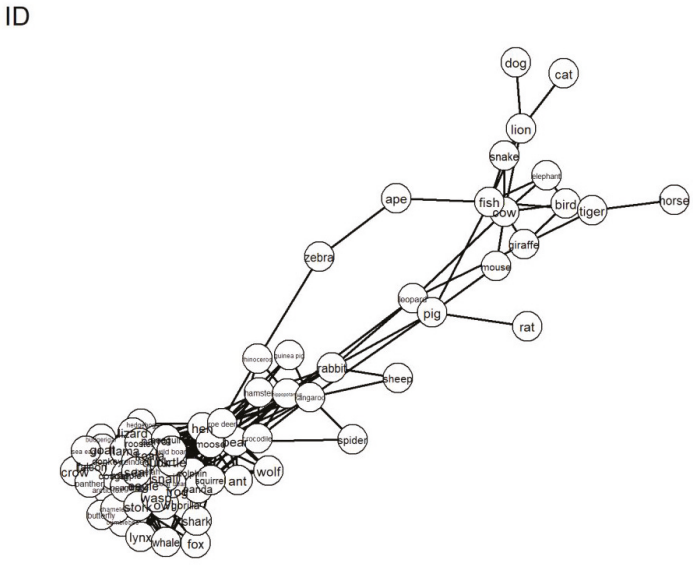


Figure 2. Graphical representation of the semantic network of the adolescents with ID.

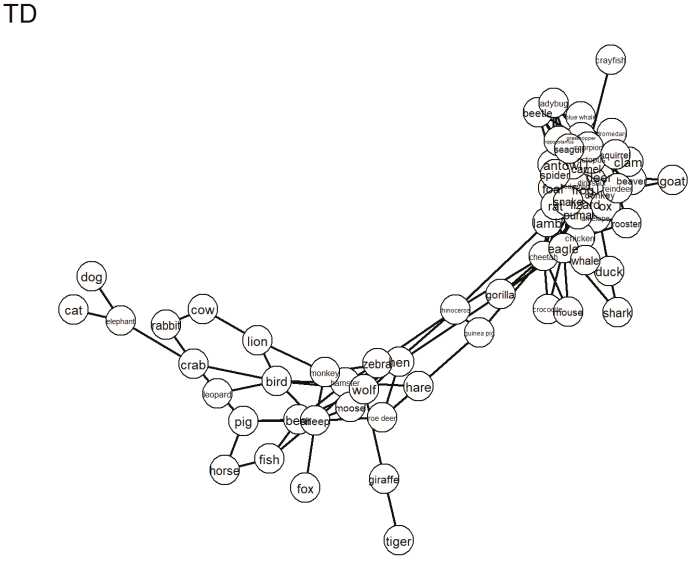


Figure 3. Graphical representation of the semantic network of the children with TD.

4. Discussion

The current study compared the semantic network structure in a group of adolescents with ID and a group of younger children with TD, matched on the produced number of words on a semantic fluency task. This is, to our knowledge, the first attempt to quantify the semantic network in the ID population. The results showed that the structure of the semantic networks differs between the groups. The semantic network of the adolescents with ID has a significantly smaller ASPL and Q and a significantly larger CC compared to with the semantic network of the children with TD. Adolescents with ID in this study

have a more condensed semantic network structure compared with children with TD, which indicates that the semantic network for the adolescents with ID is less developed. Similar results have been found for children with cochlear implants (CI; [24]) and second-language speakers [3]. Kenett et al. [24] compared a group of children with CI with a group of age-matched, typical-hearing peers. The CI group had a significantly smaller ASPL compared with the typical-hearing group. Kenett et al. [24] interpreted this result as the CI group having a less developed semantic network structure. Borodkin et al. [3] showed that second language speakers had a lexical network with a larger CC and a smaller Q in comparison with its first language equivalent. This result was interpreted as the second language speakers' network being less well-organized, as the words in the network were less likely to be grouped into identifiable subcategories [3]. Similar to the findings by Borodkin et al. [3], the current study found a lower Q value in the ID group compared with the TD group, indicating a less developed taxonomic structure of the semantic network in adolescents with ID.

Wulff et al. [1] proposed a framework for understanding the mechanisms behind age differences in the mental lexicon. We suggest that the components of this framework can be used in explaining the less developed semantic network of adolescents with ID. Wulff et al. [1] argue that the environment plays an important role in the structure of the semantic lexicon. It could be that the quality and/or quantity of the language input differs between adolescents with ID and children with TD.

The adolescents in the current study were all enrolled in special schools, meaning that they were exposed to a different learning environment compared to the children with TD. The special schools in Sweden follow a different curriculum [22]. This curriculum also provides more opportunities for individual adaptations of teaching [22], meaning that the learning environment might vary between students enrolled at the same special school. The language input for the adolescents with ID can therefore be assumed to be heterogeneous, which in turn means that a greater variation in verbal fluency performance can be expected. This could be a contributing factor as to why the estimated semantic network is less structured. Similar reasoning was used by Borodkin et al. [3], who argued that a possible explanation for the less well-organized semantic network in second language speakers could be the heterogeneous language proficiency in that group.

There has been some evidence that adolescents with ID engage in different out-of-school activities compared to their typically developing peers [23]. The difference in educational and out-of-school environments may affect the quality and/or the quantity of the linguistic input. In addition, it has been shown that parents of children with a delayed language development tend to adjust their language level on several quality measures [38], and in line with the reasoning of Beckage et al. [39], this could create a linguistic environment with different structural properties compared to the TD group.

Wulff et al. [1] proposed learning as another component that is vital for the mental lexicon. As laid out in the Introduction, statistical learning is of importance for the development of the semantic network (see: [16]). A less developed semantic lexicon for adolescents with ID could be explained by reduced statistical learning ability. In addition, an important aspect of learning is prior knowledge, meaning that the encoding of new information is moderated by pre-existing knowledge [1]. Studies have shown that the level of acquired language predicts further learning from distributional cues in infants [40,41], and suggestions have been made that the delayed language development may constrain the usage of cues [21]. Kover [20] argues that even if persons with ID may have more experience as measured by chronological time, the knowledge might be less accumulated due to poorer learning efficiency.

Currently, little is known about the effects that the structure of the semantic network has on the higher-order language ability of adolescents with ID. A less structured semantic network likely makes language understanding and production more demanding, as words might not be activated automatically (or the wrong ones might be activated). This is in

accordance with studies showing that a shorter ASPL and a higher CC might make it harder to identify words and might lead to confusing words in memory [42,43].

To conclude, adolescents with ID have a less structured semantic network than children with TD even when the network size is controlled for. These differences might be due to differences in the language environment as well as to differences in cognitive skills. If the language environment is an important factor for the structure of the semantic network of persons with ID, interventions should aim to increase the quality and quantity of the language input that children and adolescents with ID receive. The less structured semantic network might be an important underlying factor for language problems in persons with ID.

4.1. Future Studies

This is a novel field of research, and more studies are needed to disentangle the effects of different factors on the semantic network structure in persons with ID. One way of differentiating the effect of cognitive ability and the effect of the language input could be cognitive modeling. A simulation study using a semantic network model could help to investigate which type of behavior a network would display with less qualitative language input and which behavior it would display with reduced statistical learning ability. This kind of study could also help to investigate how the structure of the semantic network is influencing language ability in persons with ID. The magnitude of the differences in the current study was small (cf. [24]), and it is currently not known if these small differences in the structure influence real-life language abilities. In addition, more studies are needed to investigate the effects of different learning environments and their relation to the quality and quantity of language input.

4.2. Limitations

This study was conducted using data from two different research projects. A coordinated data collection would have allowed the research team to collect more data on related linguistic and cognitive abilities. The sample size in this study should be considered large concerning the tradition within disability research. However, when estimating networks, a larger sample size would be desirable to make sure the estimated networks are stable.

Author Contributions: Conceptualization, K.N., M.S., L.P., M.I., A.L., H.D., M.A. and D.S.; methodology, M.S., L.P. and H.D.; software, M.S., L.P. and H.D.; validation, M.S.; formal analysis, M.S., L.P. and H.D.; investigation, K.N., M.A. and M.S.; data curation, M.A. and L.P.; writing—original draft preparation, K.N., M.S., L.P., M.I., A.L., H.D., M.A. and D.S.; writing—review and editing, K.N., M.S., L.P., M.I., A.L., H.D., M.A. and D.S.; visualization, M.S.; administration, K.N., M.S. and L.P.; funding acquisition, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Union Seventh Framework Program (FP7/2007–2013) under Grant Agreement FP7-607139 (iCARE) and by the Swedish Research Council (2013-01363 and 2016-04217).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the regional Research Ethics Committee in Linköping, Sweden (2017/139-31 and 2015/308-31).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data on the children with TD presented in this study are available on request from the corresponding author. The data are not publicly available due to restrictions in the ethics application. The data for the adolescents with ID are available on https://osf.io/aet7b/ (accessed on 23 April 2021).

Acknowledgments: We would like to thank Alexander P. Christensen for the help and support with the SemNet packages in R.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Random Network comparison for the network of the children with TD, $p < 0.001$ for all comparisons.

	TD	***M***	***SD***
ASPL	3.47	2.25	0.03
CC	0.38	0.28	0.02
Q	0.35	0.23	0.01

Table A2. Random Network comparison for the network of the adolescents with ID (dataset 1), $p < 0.001$ for all comparisons.

	ID	***M***	***SD***
ASPL	2.73	2.03	0.02
CC	0.45	0.37	0.02
Q	0.31	0.17	0.01

Table A3. Random Network comparison for the network of the adolescents with ID (dataset 2), $p < 0.001$ for all comparisons.

	ID	***M***	***SD***
ASPL	2.82	2.08	0.02
CC	0.45	0.34	0.02
Q	0.33	0.19	0.01

Table A4. Random Network comparison for the network of the adolescents with ID (dataset 3), $p < 0.001$ for all comparisons.

	ID	***M***	***SD***
ASPL	3.23	2.09	0.02
CC	0.47	0.38	0.02
Q	0.33	0.19	0.01

Table A5. Random Network comparison for the network of the adolescents with ID (dataset 4), $p < 0.001$ for all comparisons.

	ID	***M***	***SD***
ASPL	2.62	2.01	0.01
CC	0.47	0.31	0.01
Q	0.36	0.18	0.01

Table A6. Random Network comparison for the network of the adolescents with ID (dataset 5), $p < 0.001$ for all comparisons.

	ID	***M***	***SD***
ASPL	2.70	1.97	0.01
CC	0.50	0.45	0.02
Q	0.28	0.15	0.01

Table A7. Random Network comparison for the network of the adolescents with ID (dataset 6), $p < 0.001$ for all comparisons.

	ID	*M*	*SD*
ASPL	2.79	1.98	0.01
CC	0.50	0.45	0.02
Q	0.23	0.14	0.01

Table A8. Random Network comparison for the network of the adolescents with ID (dataset 7), $p < 0.001$ for all comparisons.

	ID	*M*	*SD*
ASPL	3.12	2.1305	0.0177
CC	0.44	0.31	0.02
Q	0.33	0.20	0.01

Table A9. Random Network comparison for the network of the adolescents with ID (dataset 8), $p < 0.001$ for all comparisons.

	ID	*M*	*SD*
ASPL	2.98	2.05	0.02
CC	0.46	0.34	0.02
Q	0.32	0.18	0.01

Table A10. Random Network comparison for the network of the adolescents with ID (dataset 9), $p < 0.001$ for all comparisons.

	ID	*M*	*SD*
ASPL	3.12	2.04	0.02
CC	0.46	0.38	0.02
Q	0.30	0.17	0.01

Table A11. Random Network comparison for the network of the adolescents with ID (dataset 10), $p < 0.001$ for all comparisons.

	ID	*M*	*SD*
ASPL	2.57	2.08	0.02
CC	0.41	0.32	0.02
Q	0.30	0.19	0.01

References

1. Wulff, D.U.; De Deyne, S.; Jones, M.N.; Mata, R. New Perspectives on the Aging Lexicon. *Trends Cogn. Sci.* **2019**, *23*, 686–698. [CrossRef]
2. Bertola, L.; Mota, N.B.; Copelli, M.; Rivero, T.; Diniz, B.S.; Ribeiro, M.A.R.S.; Ribeiro, S.; Malloy-Diniz, L.F. Graph analysis of verbal fluency test discriminate between patients with Alzheimer's disease, Mild Cognitive Impairment and normal elderly controls. *Front. Aging Neurosci.* **2014**, *6*, 1–10. [CrossRef]
3. Borodkin, K.; Kenett, Y.N.; Faust, M.; Mashal, N. When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition* **2016**, 12. [CrossRef]
4. Brooks, P.J.; Maouene, J.; Sailor, K.; Seiger-Gardner, L. *Modeling the Semantic Networks of School-Age Children with Specific Language Impairment and Their Typical Peers*; Cascadilla Press: Somerville, MA, USA, 2017.

5. Goñi, J.; Arrondo, G.; Sepulcre, J.; Martincorena, I.; Vélez De Mendizábal, N.; Corominas-Murtra, B.; Bejarano, B.; Ardanza-Trevijano, S.; Peraita, H.; Wall, D.P.; et al. The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory. *Cogn. Process.* **2011**, *12*, 183–196. [CrossRef]
6. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*; American Psychiatric Association: Arlington, TX, USA, 2013.
7. Danielsson, H.; Henry, L.A.; Messer, D.; Ronnberg, J. Strengths and weaknesses in executive functioning in children with intellectual disability. *Res. Dev. Disabil.* **2012**, *33*, 600–607. [CrossRef]
8. Henry, L.A. The episodic buffer in children with intellectual disabilities: An exploratory study. *Res. Dev. Disabil.* **2010**, *31*, 1609–1614. [CrossRef] [PubMed]
9. van Wingerden, E.; Segers, E.; van Balkom, H.; Verhoeven, L. Cognitive and linguistic predictors of reading comprehension in children with intellectual disabilities. *Res. Dev. Disabil.* **2014**, *35*, 3139–3147. [CrossRef]
10. Henry, L.A.; Winfield, J. Working memory and educational achievement in children with intellectual disabilities. *J. Intellect. Disabil. Res.* **2010**, *54*, 354–365. [CrossRef]
11. Vitevitch, M.S. What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *J. Speech Lang. Hear. Res.* **2008**, *51*, 408–422. [CrossRef]
12. Siew, C.S.Q.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, *2019*. [CrossRef]
13. Zemla, J.C.; Austerweil, J.L. Estimating Semantic Networks of Groups and Individuals from Fluency Data. *Comput. Brain Behav.* **2018**, *1*, 36–58. [CrossRef]
14. Delis, D.C.; Kaplan, E.; Kramer, J.H. *Delis-Kaplan Executive Function System (D-KEFS)*; Psychological Corporation: London, UK, 2001.
15. Newman, M.E. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [CrossRef]
16. Unger, L.; Savic, O.; Sloutsky, V.M. Statistical regularities shape semantic organization throughout development. *Cognition* **2020**, *198*, 104190. [CrossRef] [PubMed]
17. Saffran, J.R. Statistical learning as a window into developmental disabilities. *J. Neurodev. Disord.* **2018**, *10*, 35. [CrossRef]
18. Atwell, J.A.; Conners, F.A.; Merrill, E.C. Implicit and Explicit Learning in Young Adults With Mental Retardation. *Am. J. Ment. Retard.* **2003**, *108*, 13. [CrossRef]
19. Vinter, A.; Detable, C. Implicit learning in children and adolescents with mental retardation. *Am. J. Ment. Retard.* **2003**, *108*, 94–107. [CrossRef]
20. Kover, S.T. Distributional Cues to Language Learning in Children With Intellectual Disabilities. *Lang. Speech Hear. Serv. Sch.* **2018**, *49*, 653–667. [CrossRef]
21. Thiessen, E.D.; Girard, S.; Erickson, L.C. Statistical learning and the critical period: How a continuous learning mechanism can give rise to discontinuous learning: Statistical learning and the critical period. *Wires Cogn. Sci.* **2016**, *7*, 276–288. [CrossRef]
22. Skolverket. *Läroplan för Grundsärskolan 2011: Reviderad*; Skolverket: Stockholm, Sweden, 2018.
23. King, M.; Shields, N.; Imms, C.; Black, M.; Ardern, C. Participation of children with intellectual disability compared with typically developing children. *Res. Dev. Disabil.* **2013**, *34*, 1854–1862. [CrossRef] [PubMed]
24. Kenett, Y.N.; Wechsler-Kashi, D.; Kenett, D.Y.; Schwartz, R.G.; Ben-Jacob, E.; Faust, M. Semantic organization in children with cochlear implants: Computational analysis of verbal fluency. *Front. Psychol.* **2013**, *4*, 543. [CrossRef] [PubMed]
25. Socher, M. *Reasons for Language: Language and Analogical Reasoning Ability in Children with Cochlear Implants and Children with Typical Hearing*; Linköping University Electronic Press: Linköping, Sweden, 2020; ISBN 978-91-7929-791-6.
26. Nilsson, K.; Danielsson, H.; Elwér, Å.; Messer, D.; Henry, L.A.; Samuelsson, S. Decoding Abilities in Adolescents with Intellectual Disabilities: The Contribution of Cognition, Language, and Home Literacy. *J. Cogn.* under review.
27. Socher, M.; Lyxell, B.; Ellis, R.; Gärskog, M.; Hedström, I.; Wass, M. Pragmatic Language Skills: A Comparison of Children With Cochlear Implants and Children Without Hearing Loss. *Front. Psychol.* **2019**, *10*. [CrossRef]
28. Chan, A.S.; Butters, N.; Salmon, D.P.; Johnson, S.A.; Paulsen, J.S.; Swenson, M.R. Comparison of the semantic networks in patients with dementia and amnesia. *Neuropsychology* **1995**, *9*, 177–186. [CrossRef]
29. Paulsen, J.S.; Romero, R.; Chan, A.; Davis, A.V.; Heaton, R.K.; Jeste, D.V. Impairment of the semantic network in schizophrenia. *Psychiatry Res.* **1996**, *63*, 109–121. [CrossRef]
30. R Core Team. *R: A language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
31. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
32. Wickham, H.; Bryan, J. Readxl: Read eXcel Files. Available online: https://CRAN.R-project.org/package=readxl (accessed on 27 May 2021).
33. Izrailev, S. Tictoc: Functions for Timing R Scripts, as Well as Implementations of Stack and List Structures. Available online: https://CRAN.R-project.org/package=tictoc (accessed on 27 May 2021).
34. Bååth, R. Beepr: Easily Play Notification Sounds on Any Platform. Available online: https://CRAN.R-project.org/package=beepr (accessed on 27 May 2021).

35. Wickham, H. Stringr: Simple, Consistent Wrappers for Common String Operations. Available online: https://CRAN.R-project.org/package=stringr (accessed on 27 May 2021).
36. Gohel, D. Flextable: Functions for Tabular Reporting. Available online: https://CRAN.R-project.org/package=flextable (accessed on 27 May 2021).
37. Christensen, A.; Kenett, Y. Semantic Network Analysis (SemNA): A Tutorial on Preprocessing, Estimating, and Analyzing Semantic Networks. *PsyArXiv* **2019**, 1–56. [CrossRef]
38. Vigil, D.C.; Hodges, J.; Klee, T. Quantity and quality of parental language input to late-talking toddlers during play. *Child. Lang. Teach. Ther.* **2005**, *21*, 107–122. [CrossRef]
39. Beckage, N.; Smith, L.; Hills, T. Small Worlds and Semantic Network Growth in Typical and Late Talkers. *PLoS ONE* **2011**, *6*, e19348. [CrossRef] [PubMed]
40. Lany, J. Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child. Dev.* **2014**, *85*, 1727–1739. [CrossRef] [PubMed]
41. Lany, J.; Saffran, J.R. Interactions between statistical and semantic information in infant language development. *Dev. Sci.* **2011**, *14*, 1207–1219. [CrossRef]
42. Chan, K.Y.; Vitevitch, M.S. The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **2009**, *35*, 1934–1949. [CrossRef] [PubMed]
43. Vitevitch, M.S.; Chan, K.Y.; Roodenrys, S. Complex network structure influences processing in long-term and short-term memory. *J. Mem. Lang.* **2012**, *67*, 30–44. [CrossRef] [PubMed]

big data and cognitive computing

MDPI

Article

A Novel Approach to Learning Models on EEG Data Using Graph Theory Features—A Comparative Study

Bhargav Prakash , Gautam Kumar Baboo and Veeky Baths *

Cognitive Neuroscience Lab, Department of Biological Sciences, BITS, Pilani-K.K. Birla Goa Campus, Sancoale 403726, Goa, India; f20170157@goa.bits-pilani.ac.in (B.P.); p20130404@goa.bits-pilani.ac.in (G.K.B.)
* Correspondence: veeky@goa.bits-pilani.ac.in; Tel.: +91-832-258-0436

Abstract: Brain connectivity is studied as a functionally connected network using statistical methods such as measuring correlation or covariance. The non-invasive neuroimaging techniques such as Electroencephalography (EEG) signals are converted to networks by transforming the signals into a Correlation Matrix and analyzing the resulting networks. Here, four learning models, namely, Logistic Regression, Random Forest, Support Vector Machine, and Recurrent Neural Networks (RNN), are implemented on two different types of correlation matrices: Correlation Matrix (static connectivity) and Time-resolved Correlation Matrix (dynamic connectivity), to classify them either on their psychometric assessment or the effect of therapy. These correlation matrices are different from traditional learning techniques in the sense that they incorporate theory-based graph features into the learning models, thus providing novelty to this study. The EEG data used in this study is trail-based/event-related from five different experimental paradigms, of which can be broadly classified as working memory tasks and assessment of emotional states (depression, anxiety, and stress). The classifications based on RNN provided higher accuracy (74–88%) than the other three models (50–78%). Instead of using individual graph features, a Correlation Matrix provides an initial test of the data. When compared with the Time-resolved Correlation Matrix, it offered a 4–5% higher accuracy. The Time-resolved Correlation Matrix is better suited for dynamic studies here; it provides lower accuracy when compared to the Correlation Matrix, a static feature.

Keywords: EEG; emotional states; working memory; depression; anxiety; graph theory; classification; machine learning; neural networks

Citation: Prakash, B.; Baboo, G.K.; Baths, V. A Novel Approach to Learning Models on EEG Data Using Graph Theory Features—A Comparative Study. *Big Data Cogn. Comput.* **2021**, *5*, 39. https://doi.org/10.3390/bdcc5030039

Academic Editor: Min Chen

Received: 15 June 2021
Accepted: 25 August 2021
Published: 28 August 2021

1. Introduction

Electroencephalography (EEG) is a commonly used neuroimaging tool. Its application ranges from clinical capacity such as sleep disorder studies, to seizure detection, to commercial circumstances such as EEG-controlled games [1]. The EEG data is represented as a two-dimensional matrix, which consists of electric potentials on one axis and the electrode number on the other axis. This form of EEG data makes it easy to use in machine learning models [2]. With its high temporal resolution, EEG data can provide information regarding the functional connectivity within the brain, thereby providing a topological understanding of the functioning of the human brain [3]. This is usually carried out by transforming the electrical potentials into a Correlation Matrix [4].

Functional connectivity is time dependent and to understand the functional aspects of the brain under conditions of executive functions and emotional states viz. depressive or anxious, it is vital to study them in terms of networks and the best way to do it, with the help of EEG signals, which have the highest temporal resolution in the field of neuroimaging techniques. At present learning, models use either the properties of the EEG signal such as amplitude, frequency, and event-related potentials as features or graph properties such as centrality measures which are nodal metrics or edge metrics such as shortest path length.

Network analysis and learning models on neuroimaging data have enabled researchers to study the human brain's functional and structural connectivity [5]. Here, graph metrics

are used as features for a deep learning model, apart from the standard spectral and temporal characteristics that are traditionally used [6]. Different static and dynamic features are studied to understand which features are best suited for visual working memory tasks [7]. Both Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are tested and validated for their performance on the datasets.

Previous work on emotional states such as depression and anxiety in the space of EEG and machine learning was carried out using signal features such as power or frequency bands [8]. Learning models such as probabilistic, nearest neighbor, neural network, and tree-based have been implemented on DASS scores, here the Random Forest model provided accuracy in classification of three states, i.e., depressive anxious or stressed at 84%, 85%, and 84% [9,10].

A study on clinically depressed patients and normal controls with the implementation of learning models on EEG signals using features such as frequency bands and non-linear features such as detrended fluctuation analysis (DFA), Higuchi fractal, correlation dimension, and Lyapunov exponent provided an 83.3% accuracy while using Logistic Regression [11]. Similarly, visual and verbal working memory studies using EEG have been carried out using event-related potentials (ERPs) and the subsequent construction of functional connectivity of these ERPs [5]. Studies using EEG and deep learning models involve EEG signals broken into smaller windows for training and testing [12]. The high temporal resolution being the nature of EEG signals adds an additional step in curating these smaller datasets for analysis. This step can induce a bias based on cognitive noise between participants. An SVM implementation to classify Schizophrenic patients and healthy controls based on a working memory task yields an accuracy of >74% [13].

Learning models on EEG data recorded during visual short-term memory task included SVM and Random Forest, which used raw EEG signals and the psychometric assessment scores and reaction times which provided an accuracy of approximately 90% [14]. Other implementations of SVM using frequency bands as features on similar psychological tests yield a 98% accuracy [15]. While using ERPs in the time domain, power spectra and eye-tracking as features provided accuracy in the range of 40% to 60% [7].

The intermediate step between EEG signal analysis and functional connectivity analysis is the use of a Correlation Matrix. This has been used for understanding the brain connectivity in the narrow band signals [16]. The drawback to using the matrix is that it does not address the volume conduction problem or explain the association in different frequency bands. Variations of this method have proved to be helpful to understanding the brain connectivity previously [17,18]. In this study, we explore the utility of the same along with a Time-resolved Correlation Matrix for a comparative learning model study. We compare the two correlation matrices and above all the four learning models: Logistic Regression, Random Forest, Support Vector Machine, and Recurrent Neural Networks, which can shed some light on the nature of EEG activity in these emotional and cognitive states. Using the Correlation Matrix provides a non-directed graph. These kinds of graphs aim to understand the interaction between signals. This enables one to understand the dominant influence at a specific time in the brain signals [19]. Here the EEG data for working memory and emotional states are used from a total of 359 (25(DASS21), 122 (Selection Task), 29 (WM-Lab), 27 (Visual-WM+drug), and 156 (Verbal-WM)) participants. Both EEG data and associated psychometric assessment scores are used for the learning model study. Two high-accuracy models, i.e., recurrent neural network and Random Forest, belong to the neural networks method and ensemble methods. Furthermore, two high interpretable models, a kernel-based method- Support Vector Machine and the Logistic Regression model, are examined and compared.

2. Materials and Methods

2.1. Datasets

In this study, five EEG datasets are used, of which two were recorded in house, and three are from a public database. Among the two recorded in-house, 25 partici-

pants are from Sternberg Visual Working Memory Task, and 29 participants are from the DASS 21 questionnaire (Figure 1) (approved by the Institute Research Ethics Committee (IHEC-40/16-1)) using a 32 Channel(bipolar montage) EGI geodesic system (Appendix A Figure A1). From the OpenNeuro dataset, 122 participants from Probabilistic Selection Task (OpenNeuro Dataset Accession Number: ds003474) is recorded using a 64 channel Synamps system, 156 participants from verbal working memory Task (OpenNeuro Dataset Accession Number: ds003565) is recorded using a 19 channel 10-20 system Mitsar-EEG-202 amplifier, and 27 participants from visual working memory task (OpenNeuro Dataset Accession Number: ds003519) are used. A total of 359 participants' EEG data is used here (Table 1).

No.	Question	Never	Sometimes	Often	Almost Always		D	A	S
1	I found it hard to wind down	0	1	2	3				
2	I was aware of dryness of my mouth	0	1	2	3				
3	I couldn't seem to experience any positive feeling at all	0	1	2	3				
4	I experienced breathing difficulty (eg. Excessively rapid breathing, breathlessness in the absence of physical exertion)	0	1	2	3				
5	I found it difficult to work up the initiative to do things	0	1	2	3				
6	I tended to over-react to situations	0	1	2	3				
7	I experienced trembling (e.g. in the hands)	0	1	2	3				
8	I felt that I was using a lot of nervous energy	0	1	2	3				
9	I was worried about situation in which I might panic and make a fool of myself	0	1	2	3				
10	I felt that I had nothing to look forward to	0	1	2	3				
11	I found myself getting agitated	0	1	2	3				
12	I found it difficult to relax	0	1	2	3				
13	I felt down-hearted and blue	0	1	2	3				
14	I was intolerant of anything that kept me from getting on with what I was doing	0	1	2	3				
15	I felt I was close to panic	0	1	2	3				
16	I was unable to become enthusiastic about anything	0	1	2	3				
17	I felt I wasn't worth much as a person	0	1	2	3				
18	I felt that I was rather touchy	0	1	2	3				
19	I was aware of the action of my heart in the absence of physical exertion (ex. sense of heart rate)	0	1	2	3				
20	I felt scared without any good reason	0	1	2	3				
21	I felt that life was meaningless	0	1	2	3				
					Total				

Figure 1. DASS 21 questionnaire example.

DASS 21 questionnaire is a 21 item self-administered test; this test contains seven sets of questions to assess the three emotional states; depressive, anxious, and stressed. A participant responds with a score ranging from 0 to 3, with 0 meaning never and three meaning almost always. Scores for each category are cumulative; a rating between normal to severe is provided at the end of the test. These scores are then used in classifying participants for the training dataset (Figure 1). Preprocessing of the EEG files are carried out on EEGLab toolbox [20] on MATLAB. Here the data is filtered using the Basic filter option; this option uses the "pop_eegfiltnew()" function from MATLAB. The function filters the data using Hamming windowed sinc FIR filter. The filter order/transition band width is estimated with the following heuristic in default mode: transition bandwidth is 25% of the lower passband edge, but not lower than 2 Hz, where possible (for bandpass,

highpass, and bandstop) and distance from passband edge to critical frequency (DC, Nyquist) otherwise. Window type is hardcoded to Hamming. Furthermore, decomposition of data using Independent Component Analysis (ICA) [21], the filtering and ICA is carried out on the MARA toolbox [22], where the option of automatic removal of components is selected. Following which the data is exported as .set files.

Table 1. Overview of EEG datasets.

Sl. No.	Name of the Dataset	EEG Recording System	Acquisition Parameters
1	Visual Working Memory (n = 25)	32 Channel EGI geodesic	impedance < 50 kΩ, 1000 Hz sampling rate, band-pass filter 0.1–70 Hz, 50 Hz notch filter
2	Visual Working Memory (n = 27)	64-channel Brain Vision system	500 Hz sampling rate, Band-pass filter 0.1–100 Hz
3	DASS 21 Questionnaire (n = 29)	32 Channel EGI geodesic	impedance < 50 kΩ, 250 Hz sampling rate, band-pass filter 0.1–70 Hz, 50 Hz notch filter
4	Probabilistic Selection and Depression (n = 122)	64 Ag/AgCl electrodes Synamps2 system	impedance < 10 kΩ, 500 Hz sampling rate, band-pass filter 0.5–100 Hz
5	Verbal Working Memory (n = 156)	19 electrodes 10–20 system Mitsar-EEG-202 amplifier	500 Hz sampling rate, band-pass filter 1–150 Hz 50 Hz notch filter

The task, Probabilistic Selection and Depression (public database), has two tests, the Becks Depression Inventory and the State-Trait Anxiety Inventory [23]. The scores of these tests again range from normal to severe. For the Probabilistic Selection Task [24] the participants were administered the Beck Depression Inventory (BDI) and State-Trait Anxiety Inventory (STAI). Here, BDI scores that are less than or equal to 19 are considered zero and greater than or equal to 20 as one; likewise, for STAI scores, equal to and lesser than 55 are considered as zero and greater than or equal to 56 as one. Bad channels and bad epochs were identified using a conjunction of the FASTER algorithm [25] and the pop_rejchan from EEGLab [20] toolbox and were subsequently interpolated and rejected, respectively. The FASTER algorithm has epoch sensitivity of 97.54%, removes 3.1% of the epochs, and has eye blinks sensitivity of 99.07%. Eye blinks were removed following ICA. Data were re-referenced to averaged mastoids.

Visual working memory (in-house recording) is a modified Sternberg working memory task (Designs, 2021), which involves a visual chart that needs to be memorized/committed to memory, followed by tasks to complete based on the recollection of the chart from memory. Preprocessing of the EEG files are carried out on EEGLab toolbox [20] on MATLAB. Here the data is filtered using the Basic filter option. Decomposition of data using ICA [21] is performed, following which the data is exported as .set files. The recollection of the participant is tested by presenting seven questions on the basis of the visual chart: a score of 50% or less is considered as zero and above as one.

Visual Working Memory and Cabergoline (1.25 mg) Challenge [26], here a drug that can improve memory functions and placebo, is administered to a small group of participants. The placebo and drug groups are used for classification. For the Visual Working Memory and Cabergoline [27] challenge data [28], two sessions are carried out for each participant, one with a placebo and the other with the drug. Here the placebo is treated as zero and drug administered session as one. Data was visually inspected for bad

channels to be interpolated and bad epochs to be rejected. Time-frequency measures were computed by multiplying fast Fourier transformed (FFT) power spectrum of single trial EEG data with the FFT power spectrum of complex Morlet wavelets. The end result of this process is the same as time-domain signal convolution.

Finally, verbal working memory (public database) [29] consists of the EEG recorded in a modified Sternberg working memory paradigm with two types of tasks, with mental manipulations (alphabetization), simple retention (TASK), and three levels of load and 5, 6, or 7 letters to memorize (LOAD). When the participant is able to answer greater than 50% of the time in the trial it is considered one and below 50% is considered zero. First, ocular activity artifacts were addressed using ICA using AMICA algorithm [30]. Second, epochs containing artifacts were visually identified and discarded [31]. EEGLab [20] was used for data preprocessing.

Apart from exploring the utility of the Correlation Matrix, a comparison between the data recorded in-house and the public database is carried out using the accuracy of the models. EEG artifacts suppression and removal was conducted in the following two steps.

2.2. Computation of Correlation Matrix and Time Resolved Correlation Matrix Using Brainstorm Toolbox

The .set files are then imported onto the Brainstorm toolbox [32]. Here, using the Editor pipeline, the connectivity option is used for computing the Correlation Matrix and Time-resolved Correlation Matrix.

In this connectivity analysis, the following points are considered:

- The EEG sensors data is used from each of the datasets.
- Trail based data is drawn on.
- Full networks are calculated.
- In terms of temporal resolution, both static and dynamic are studied.
- The output data has a 4-D structure: Channels X Channels X Frequency Bands X Time.

2.2.1. Correlation Matrix Computation

The editor pipeline for computing the Correlation Matrix, the connectivity option for Correlation Matrix, provides three option windows as follows: Input options, Process options, and finally Output options. These options are presented in a nutshell below:

1. The Input option has three input fields, namely:
   (a) Time window.
   (b) Sensors types or names.
   (c) Checkbox to include bad channels.
2. The process option has a checkbox to allow for computing the scalar product instead of correlation.
3. Finally, output options, which has two checkboxes: (1) for saving individuals' results (one file per input file) and (2) for saving the average connectivity matrix (one file).

2.2.2. Time Resolved Matrix Computation

In case of Time resolved matrix, the editor has two main options: Input options and Process Options which are described briefly below.

1. Input option has three input fields:
   (a) Time window.
   (b) Sensor types or names and a checkbox to include bad channels.
2. Process option has:
   - Estimation window length (350 ms).
   - Sliding window overlap (50%).
   - Estimator options: computing the scalar product instead of correlation.
   - Output configuration (enables addition of comment tag).

2.3. Methods

2.3.1. Data Processing

Given the sensitivity of the EEG signals, it is imperative to preprocess them before any other analysis of the data is carried out. Therefore, the EEG data is filtered to remove line noise (50 Hz), band-pass filters, removal of bad channels, and artifact removal (please refer to Section 4.1 datasets for details), and this data is converted into a Correlation Matrix (NxN), each matrix corresponds to each EEG session file and time resolved Correlation Matrix (NxN), with a 50% sliding window overlap and 350 ms window length. The matrix is square and symmetrical, where each cell entry is the correlation between any two EEG electrodes; these operations are carried out on the Brainstorm package [33] on MATLAB.

Principal Component Analysis (PCA) is carried out with the help of scikit-learn [34] before using the data as input. This helped in two ways: it reduced the dimension of data while preserving the features and is a standard method for removing multicollinearity.

2.4. Learning Models

After preprocessing and feature extraction of the original EEG data, the Correlation Matrix (feature) is used as input to different classifiers, including traditional machine learning algorithms and neural networks tuned in line with our data. The models used are Logistic Regression, Random Forest, Support Vector Machine, and Recurrent Neural Networks (RNN) to classify the EEG data. The performance evaluation of the different classifiers is examined using a confusion matrix, whose components are T.P., TN, F.P., and F.N. Further, the accuracies are calculated using these measures, using the formula:

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) * 100 \quad (1)$$

T.P.: True Positives T.N.: True Negatives F.P.: False Positives F.N.: False Negatives.

Overfitting/underfitting: In this study the problem of overfitting did not pose an obstacle in these datasets. Hence, the results were not unusually accurate. Regularisation factors are applied to reduce overfitting. Finally for underfitting, varying the hyperparameters over a large range is carried out and the best fitting set of values is appropriated.

2.4.1. The Logistic Regression Model (LR)

A Logistic Regression model with Gaussian Kernel and Laplacian Prior is used for classification. The Gaussian kernel optimizes the separation between data points in the transformed space obtained in preprocessing, while the Laplacian Prior enhances the sparseness of learned L.R. regressors to avoid overfitting [8]. A multinomial L.R. model where the probability that an input feature xi belongs to class k is given by:

$$p(y_i = k|x_i, w) = \frac{exp(w^{(k)}h(x_i))}{\sum_{k=1}^{K} exp(w^{(k)}h(w_i)}, \quad (2)$$

x_i: feature vector
k: class
$h(x_i)$: linear transformation function of x_i
w: logistic regressors.

2.4.2. Support Vector Machine (SVM)

Apart from the application of SVM on EEG data, implementation of SVM on MRI data to classify between major depressive disorder and bipolar disorder provided accuracy up to 45% to 90% [35]. The main reason behind using SVM is to leverage its relatively less computational power to produce a significant accuracy and to reduce possible redundant information (which is very common in EEG datasets) residing in the data. The input data is mapped to a higher dimensional vector space using a linear kernel function to find a hyperplane for classification (Figure 2).

$$w * z - b = 0 \tag{3}$$

w: normal vector
b: bias of separation of hyperplane.

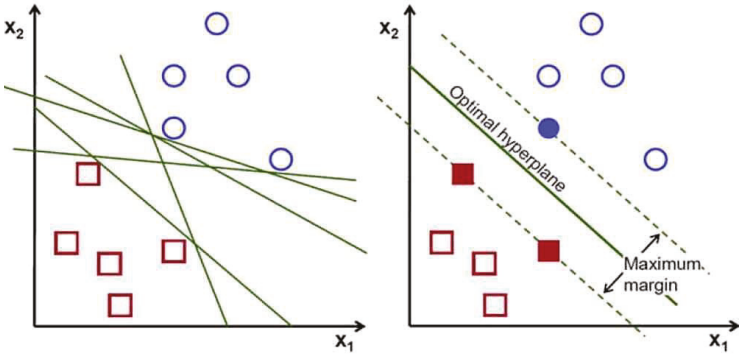


Figure 2. Support Vector Machine algorithm with the construction of different hyperplanes that separates the different classes. The most optimal hyperplane is the one that maximizes this separation.

2.4.3. Random Forest (RF)

A Random Forest classifier (Figure 3) that uses an ensemble learning approach towards prediction is used. R.F. classifier works in a similar way as the decision tree classifier, only with an ensemble learning approach added to it. The first step is the creation of many random decision trees, each predicting a particular class according to the features given to it. Once each tree predicts a class, voting is carried out to take into consideration the final class according to a majority. The output is then the class that has the majority voting.

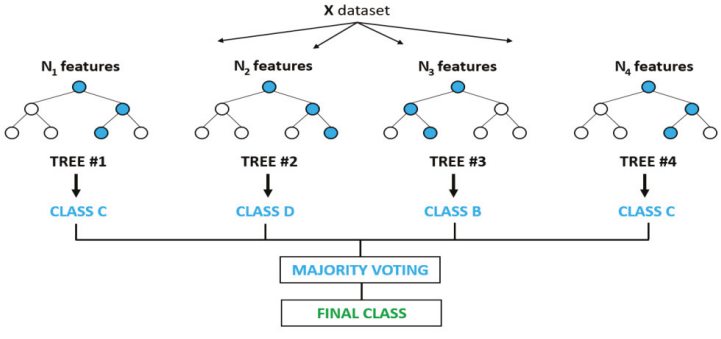


Figure 3. Ensemble method implemented by the Random Forest Algorithm. The ensemble consists of different trees fitted on the data with a range of hyperparameters. The tree which fits the data most optimally is then chosen by the algorithm by majority voting method.

2.4.4. Recurrent Neural Network (RNN)

Previous work on the implementation of neural networks on EEG signals has been fruitful, which provided accuracy in the range of 81% to 94% [36]. RNN was a good model for studying both working memory [37,38] and emotional state [39] EEG data when compared to other models such as SVM or deep belief networks [40]; on that note the following RNN model is implemented. The RNN is implemented through a Long Short Term Memory (LSTM) model [6,41], producing exemplary results on sequential data, such

as EEG data. A sequential model is used to build the LSTM, which is a linear stack of layers. The first layer is an LSTM layer with 256 memory units, and it defines the input shape. This is done to ensure that the next LSTM layer receives sequences and not just randomly scattered data. The next layer is a Dense layer with a "sigmoid" activation function. A dropout layer is applied after each LSTM layer to avoid overfitting of the model. The model is then trained and monitored for validation accuracy using loss as "binary cross-entropy", optimizer as "adam", and metrics as "accuracy" (Figure 4).

$$H(q) = -1/N \sum y_i * log(p(y_i)) + (1 - y_i) * log(1 - p(y_i)) \quad (4)$$

$H(q)$: binary cross entropy
$p(y_i)$: probability of belonging to class y_i.

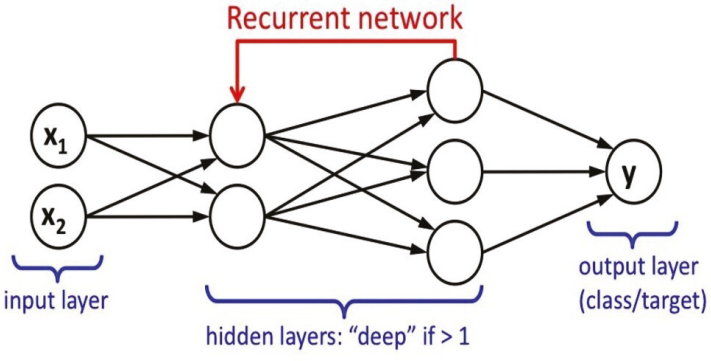


Figure 4. Recurrent Neural Network (RNN), representing the skeleton on which every RNN is built. The output of each layer acts as the input to the next and modifies the hyperparameters of the layer in each epoch, thus implementing the learning part of the algorithm.

3. Results

The performance of RNN classifiers shows up to 94.50% and 88.64% accuracies for each of the working memory tasks, which outperforms most of the previous works reviewed. The performance of R.F. and L.R. classifiers are relatively sub-par compared to RNNs but still comparable to previously obtained results. The poor performance of SVMs highlights the shortcomings of the method adopted in this study in algorithms that are sensitive to the dimensions of the data. The impressive performance of RNNs can be attributed to their innate ability to extract correlated features, which are not visible in traditional statistical methods, within the data with the help of their stacked networks and activation functions. The standard performance of R.F. and L.R. algorithms highlights the validity of the method adopted in this study and the enormous scope it provides for further improvement.

Further, the data from the public database provides higher accuracy (Tables 2 and A2) in all four models when compared to the in-house data (Tables 3 and A3). On average there seems to be a difference of 40–60% accuracy between the two groups.

Table 2. Accuracy of Classifying Emotional States from the Probabilistic Selection Task Data.

Emotional State/ Learning Model Accuracy	Logistic Regression	Random Forest	SVM	RNN
Depression	71.33%	73.46%	61.78%	88.64%
Anxiety	64.56%	78.66%	65.27%	80.75%

Table 3. Classifying Emotional States from the DASS 21 Data.

Emotional State/ (% Accuracy of Model)	Logistic Regression	Random Forest	SVM	RNN
Depression	35.06%	28.60%	27.60%	34.75%
Anxiety	28.40%	34.45%	30.85%	38.85%
Stress	31.10%	33.20%	31.70%	36.40%

4. Discussion

The current adaptation of learning models for studying brain connectivity with EEG dataset involves feature extraction from the signal itself, such as the power spectral density or event-related potentials (ERP). Or, when the EEG data is transformed into a graph, graph features such as nodal/edge metrics need to be calculated first before the machine learning process (which is done for some of the datasets that were obtained from OpenNeuro [26,42]). These steps require a dossier containing the experimental paradigms, the brain regions involved in the testing condition, specifics of band frequencies, Transition Frequency, and ERPs. This entails added/newer steps to various stages of the data processing. This translates to increased time and not to mention the myriad of statistical analyses that need to be carried out.

Methods such as phase coherence, phase locking value, or pairwise phase consistency which transform the EEG data to a matrix form for network construction, require adding steps to the analysis which translates to more time spent [19]. These methods and the convenience or inconvenience a method can add to the analysis pipeline takes the analysts on a puzzling path to addressing them with more tools to appreciate or tackle the unexpected observations or results.

Functional Connectivity using EEG data can be done on the basis of frequency, time domain, or phase characteristics of the signal. This can further be categorized as static or dynamic. The various methods under each of these categories have their own advantages and disadvantages, followed by tools/methods that can strengthen or weaken the said methods of functional connectivity analysis. Some of the common challenges with EEG-functional connectivity studies include (1) The common Reference Problem: the use of unipolar reference scheme tends to provide false coherence whereas (2) the bipolar reference scheme or (3) unipolar with separate reference address the problem inherent with unipolar reference schemes [19] (note: The EEG datasets used in this study used bipolar reference schemes).

To address the obstacle of the signal to noise (SNR) ratio, the impedance during the recordings is maintained and monitored as stringently as possible. Although, the best practice to address the SNR problem would be to use stratification methods or a suitable post-hoc method. The sample bias problem is addressed by using each of the trials as an input to the learning models. To circle back, in this study the main aim is to evaluate the performance of the four learning models.

To circle back, in this study the main aim is to evaluate the performance of the four learning models.

Implementation of learning models on imaging data to study emotional states provided reliable results in the past [43]. With the use of both high accuracy (RNN and R.F.) and high interpretability (SVM and L.R. model), we can look for non-linear relationships, non-smooth relationships, and well-defined relationships.

Comparison of learning models on similar paradigm EEG data helps with functional connectivity study. Here, it is demonstrated that a Correlation Matrix can be used in learning models and provides exemplary accuracy. Furthermore, it yielded higher accuracy rates with well-structured data obtained in a controlled environment, as with the working memory tasks, indicating superior discriminatory performances when assessing mental tasks. In addition, the present study is discriminatory towards poorly collected and insufficient data.

From running the classification models on both types of datasets: correlation and Time-resolved Correlation Matrices, we find that the two classification models: Random Forest Classifier and RNN classifier, perform relatively better when the correlation is not time-resolved. The performance dips across both the Verbal Memory and Working Memory datasets for time-resolved correlation. This provides scope for further research as to why dynamic methods may not be a better fit for Neural Networks and Decision Tree based classification models.

This study sheds some light on brain networks when studying emotion or executive functions such as working memory. By using the Correlation Matrix as such, this enables us to study the brain activity as a complete network and not sub-networks or brain regions [44]. Another underlying quality of the participants is their linguistic abilities. The data collected in-house had participants who were at least bilingual, of the 59 participants only 5 were bilinguals and the rest either had adequate knowledge of a third language or fourth. Similarly, the datasets of the probabilistic selection task [45] and the verbal working memory task [42] consist of participants who know English as well as Japanese (selection task) and Russian (working memory). The results from this study and the need to understand the bilingual/multilingual neurocognition [46] of individuals necessitate a deeper study into the role of language on emotional states and working memory. A comprehensive study into the static/dynamic metrics under the three categories of time, frequency, and phase would help in understanding which methods and which parameters can work best for a particular experimental paradigm.

The engagement of participants in the five experimental paradigms is by nature dynamic states of EEG activity. With this in mind, the use of a Time-resolved Correlation Matrix is explored alongside the Correlation Matrix, both features of the time domain of the signal. Since the results indicate the use of the dynamic feature is most suitable for such cognitive states, it sets the stage to explore the other static and dynamic features of trial-related EEG data. At this stage, this investigation provides a step for exploring the possibility of using these features as markers for the cognitive footprints of psychopathologies such as memory and emotional state deficits.

The results indicate that using graph metric for dynamic (Correlation Matrix) features is optimum. Computerized administration of the test rules out pressure to perform or dishonesty.

4.1. Limitations

Given that both the positive and negative lag indicates an influence in the network, the bi-directional interactions that could be occurring are beyond the scope of the current study [19]. In this regard, it is to be noted that in EEGLab, filtering for connectivity analysis can be carried out using the "Basic Filter (legacy)". This applies the filter forward and then again backward to ensure that phase delays introduced by the filter are nullified. The "causal filter" (part of the "Basic Filter (legacy)") is only applied forward, so phase delays might be introduced and not compensated for. However, causal relationships are preserved. This introduces the problem of phase distortion. In this study, the common input problem is not dealt with as it would increase the number of steps involved in the preprocessing of the EEG data and also increase the run time of the pipeline. The current study does not have the resolution to examine the salience, executive, and task-related networks or provide a distinction between the three [44].

Although RNN and Random Forest models provide high accuracy, both these methods have longer run times when compared to the other two. In the current study, the lack of defined healthy control groups across the datasets can be addressed, which can help improve the accuracy of the models. This imbalance can be addressed using larger data and a robust learning model [47].

Single trials in the case of the in-house dataset and using DASS 21 for the first time as a computerized test and EEG could explain the lower accuracy across the models associated with this data. This also applies to the visual working memory data recorded in the lab.

The EEG data is collected from four different EEG acquisition systems with five different acquisition parameters. Furthermore, the experimental paradigms are dissimilar along with the distribution of participants among the two main study areas, i.e., emotional states (n = 176) and working memory (n = 183) which is uneven.

Using graph features on the EEG data is time consuming because graph features can range from nodal metrics to local/global network characteristics that need to be considered features. Simultaneously cherry-picking graph metric(s) can introduce a bias that has to be considered in the study and addressed at a later point with defined statistical analysis.

5. Conclusions

The time-series nature of the EEG data, which is an effective form of neuroimaging data for studying the functional connectivity of the brain, is studied for its utility in a machine learning environment. Although this is not a first of its kind, the use of the Correlation Matrix/Time-resolved Correlation Matrix makes it one. The previous work on implementing learning models on EEG data consists of using features from the signal processing field. These studies provide insight into the possible electrical activity of each lobe(s) associated with the behavior. However, they fall short while explaining the possible functional connectivity between the regions of the brain or the whole brain. Using such EEG datasets recorded on the working memory and emotional state assessment paradigms, a preliminary comparison of the different EEG acquisition systems and acquisition parameters is attempted.

The application of the Correlation Matrix can be implemented as a first step into choosing the appropriate learning model for studying the emotional or working memory EEG data. This study reveals that using a Correlation Matrix instead of a Time-resolved Correlation Matrix even under trail-based EEG data is a better-suited input for learning models when compared to a dynamic feature such as the Time-resolved Correlation Matrix. This brings us to the experiments themselves.

The memory tasks and psychometric assessment tests—BDI, STAI, and DASS 21—involve different brain regions, given that they have to be functionally connected to respond to the questions in these tests. This study provides a basis for studying the cognitive footprints for memory deficits, depression, anxiety, and stress. Further, it is observed that RNN performs the best compared with the other three models implemented in this study.

Author Contributions: Conceptualization, G.K.B. and V.B.; methodology, B.P.; software, B.P.; validation, G.K.B. and B.P.; formal analysis, G.K.B. and B.P.; investigation, G.K.B. and B.P.; resources, V.B.; writing—original draft preparation, G.K.B. and B.P.; writing—review and editing, V.B.; visualization, B.P. and G.K.B.; supervision, V.B.; project administration, V.B.; funding acquisition, Veeky Baths. All authors have read and agreed to the published version of the manuscript.

Funding: We thank the Department of Science and Technology, Government of India for the grant (SR/CSRI/50/2014(G)) and Department of Biological Sciences, BITS, Pilani-K.K. Birla Goa Campus for the infrastructure support.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of Birla Institute of Technology and Science, Pilani (IHEC-40/16-1).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study. Written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: The data presented in this study are openly available in OpenNEURO repository-

- EEG: Visual Working Memory + Carbergoline Challenge Dataset
  https://openneuro.org/datasets/ds003519/versions/1.1.0
  DOI:10.18112/openneuro.ds003519.v1.1.0,
- EEG: Probabilistic selection task and Depression Dataset
  https://openneuro.org/datasets/ds003474/versions/1.1.0
  DOI:10.18112/openneuro.ds003474.v1.1.0,
- VerbalWorkingMemory Dataset
  https://openneuro.org/datasets/ds003655/versions/1.0.0
  DOI:10.18112/openneuro.ds003655.v1.0.0.

And In-house datasets can be accessed here

- DASS 21 Questionnaire EEG recordings-https://tinyurl.com/cvd729p8 and
- Working Memory EEG recordings-https://tinyurl.com/2z6ms7p6

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Reproduction of the Research Shown

1. in-house EEG datasets please follow the steps provided below
   - import the files to EEGLab on MATLAB
   - filter the files using the MARA Toolbox using band-pass filter 0.1–70 Hz and 50 Hz notch filter
   - please select automatic ICA rejection
   - export the files as .set format
   - import the files (create a study for each dataset) on to brainstorm toolbox on MATLAB.
   - use the connectivity editor for computing Correlation Matrix
2. For the OpenNEURO datasets
   - import the files(create a suitable study protocol for each dataset) on to brainstorm
   - use the connectivity editor for computing Correlation Matrix
   - In case the files on OpenNEURO are RAW files, follow the steps provided on the readme file for preprocessing of the EEG recordings.

Please follow the link-https://github.com/bhargavPrak/eeg_classification, accessed on GitHub (accessed on 26 August 2021) for brief description of the skeleton of the machine learning models implemented in this study.

Note: Please refer to the articles for each of the OpenNEURO datasets, since each of them have implemented specific and distinct preprocessing techniques which were best suited for the experimental paradigm. Any deviation from the methods used would impact the overall accuracy obtained from the machine learning models.

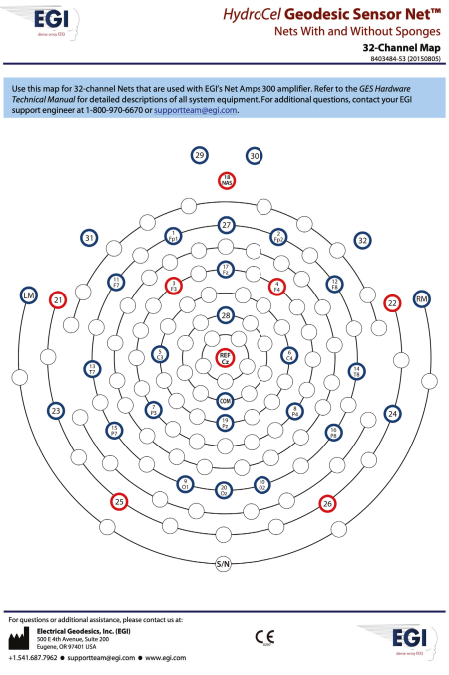


Figure A1. EEG Sensor Placement, the signals of each sensor helps in studying the activity of the particular region of the brain. This further helps in functional connectivity studies of the brain.

Table A1. Accuracy of Classifying Placebo vs. Drug induced Memory Task conditions.

Condition	Logistic Regression (% Accuracy)	Random Forest (% Accuracy)	SVM (% Accuracy)	RNN (% Accuracy)
Placebo	73.60	80.40	73.50	90.20
Drug	71.80	81.60	76.80	92.80

Table A2. Accuracy of Classifying Verbal Memory Task Conditions 5, 6 or 7 letters.

	5	6	7
Manipulation	Logistic regression–66.66% Random forest–65.50% SVM–60.15% RNN–75.86%	Logistic regression–59.40% Random forest–69.40% SVM–59.80% RNN–70.40%	Logistic regression–61.10% Random forest–76.70% SVM–54.70.10% RNN–71.50%
Retention	Logistic regression–68.70% Random forest–70.60% SVM–55.60% RNN–74.80%	Logistic regression–66.40% Random forest–65.80% SVM–50.20% RNN–70.60%	Logistic regression–63.40% Random forest–68.30% SVM–53.30% RNN–79.60%

Table A3. Participants of Modified Sternberg Working Memory Task.

	Logistic Regression (% Accuracy)	Random Forest (% Accuracy)	SVM (% Accuracy)	RNN (% Accuracy)
Participant 01	12.5	37.5	28.60	12.5
Participant 02	25	28.30	28.60	28.60
Participant 03	14.30	37.5	14.30	14.30
Participant 04	50	12.5	25	25
Participant 05	25	25	25	28.60
Participant 06	25	12.5	12.5	14.30
Participant 07	14.30	42.90	12.5	50
Participant 08	12.5	25	12.5	12.5
Participant 09	50	28.60	22.22	25
Participant 10	75	50	14.60	14.60
Participant 11	12.5	12.5	28.60	22.22
Participant 12	37.5	50	11.11	12.5
Participant 13	28.60	14.30	25	28.60
Participant 14	12.5	12.5	37.5	14.30
Participant 15	25	25	37.5	25
Participant 16	25	12.5	12.5	12.5
Participant 17	28.60	25	50	33.33
Participant 18	12.5	37.5	25	14.60
Participant 19	50	12.5	37.5	25
Participant 20	14.30	14.30	14.30	12.5
Participant 21	25	37.5	14.30	12.5
Participant 22	12.5	25	22.22	14.30
Participant 23	14.30	25	28.60	25
Participant 24	25	12.5	12.5	28.60
Participant 25	50	28.60	12.5	12.5

References

1. Soufineyestani, M.; Dowling, D.; Khan, A. Electroencephalography (EEG) technology applications and available devices. *Appl. Sci.* **2020**, *10*, 7453. [CrossRef]
2. Li, G.; Lee, C.H.; Jung, J.J.; Youn, Y.C.; Camacho, D. Deep learning for EEG data analytics: A survey. In *Concurrency Computation*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2019. [CrossRef]
3. Vecchio, F.; Miraglia, F.; Maria Rossini, P. Connectome: Graph theory application in functional brain network architecture. *Clin. Neurophysiol. Pract.* **2017**, *2*, 206–213. [CrossRef] [PubMed]
4. Wendling, F.; Ansari-Asl, K.; Bartolomei, F.; Senhadji, L. From EEG signals to brain connectivity: A model-based evaluation of interdependence measures. *J. Neurosci. Methods* **2009**, *183*, 9–18. [CrossRef] [PubMed]
5. Bashiri, M.; Mumtaz, W.; Malik, A.S.; Waqar, K. EEG-based brain connectivity analysis of working memory and attention. In Proceedings of the ISSBES 2015-IEEE Student Symposium in Biomedical Engineering and Sciences: By the Student for the Student, Shah Alam, Malaysia, 4 November 2015; pp. 41–45. [CrossRef]
6. Chang, S.; Dong, W.; Jun, H. Use of electroencephalogram and long short-term memory networks to recognize design preferences of users toward architectural design alternatives. *J. Comput. Des. Eng.* **2020**, *7*, 551–562. [CrossRef]
7. Krumpe, T.; Scharinger, C.; Rosenstiel, W.; Gerjets, P.; Spüler, M. Unity and diversity in working memory load: Evidence for the separability of the executive functions updating and inhibition using machine learning. *bioRxiv* **2018**. [CrossRef]
8. Wu, C.T.; Dillon, D.; Hsu, H.C.; Huang, S.; Barrick, E.; Liu, Y.H. Depression Detection Using Relative EEG Power Induced by Emotionally Positive Images and a Conformal Kernel Support Vector Machine. *Appl. Sci.* **2018**, *8*, 1244. [CrossRef]
9. Kumar, P.; Garg, S.; Garg, A. Assessment of Anxiety, Depression and Stress using Machine Learning Models. *Procedia Comput. Sci.* **2020**, *171*, 1989–1998. [CrossRef]
10. Priya, A.; Garg, S.; Tigga, N.P. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Comput. Sci.* **2020**, *167*, 1258–1267. [CrossRef]
11. Hosseinifard, B.; Moradi, M.H.; Rostami, R. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput. Methods Programs Biomed.* **2013**, *109*, 339–345. [CrossRef]
12. Schirrmeister, R.; Gemein, L.; Eggensperger, K.; Hutter, F.; Ball, T. Deep learning with convolutional neural networks for decoding and visualization of eeg pathology. *arXiv* **2017**, arXiv:1708.08012.
13. Johannesen, J.K.; Bi, J.; Jiang, R.; Kenney, J.G.; Chen, C.M.A. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatr. Electrophysiol.* **2016**, *2*, 1–21. [CrossRef]

14. Antonijevic, M.; Zivkovic, M.; Arsic, S.; Jevremovic, A. Using AI-Based Classification Techniques to Process EEG Data Collected during the Visual Short-Term Memory Assessment. *J. Sens.* **2020**, *2020*, 8767865. [CrossRef]
15. Amin, H.U.; Mumtaz, W.; Subhani, A.R.; Saad, M.N.M.; Malik, A.S. Classification of EEG signals based on pattern recognition approach. *Front. Comput. Neurosci.* **2017**, *11*, 103. [CrossRef]
16. Ruiz-Gómez, S.J.; Hornero, R.; Poza, J.; Santamar'ia-Vázquez, E.; Rodr'iguez-González, V.; Maturana-Candelas, A.; Gómez, C. A new method to build multiplex networks using canonical correlation analysis for the characterization of the Alzheimer's disease continuum. *J. Neural Eng.* **2021**, *18*, 26002. [CrossRef] [PubMed]
17. Tanaka, H.; Miyakoshi, M. Cross-correlation task-related component analysis (xTRCA) for enhancing evoked and induced responses of event-related potentials. *NeuroImage* **2019**, *197*, 177–190. [CrossRef] [PubMed]
18. Perinelli, A.; Chiari, D.E.; Ricci, L. Correlation in brain networks at different time scale resolution. *Chaos* **2018**, *28*, 063127. [CrossRef] [PubMed]
19. Bastos, A.M.; Schoffelen, J.M. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* **2016**, *9*, 175. [CrossRef]
20. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef] [PubMed]
21. Anthony J.B.; Makeig, S.T.P.J.T.J.S. Independent Componenet Analysis of Electroencephalographic Data. *Adv. Neural Inf. Process. Syst.* **1996**, *91*, 145–151.
22. Winkler, I.; Haufe, S.; Tangermann, M. Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behav. Brain Funct.* **2011**, *7*, 30. [CrossRef] [PubMed]
23. Julian, L.J. Measures of Anxiety. *Arthritis Care* **2011**, *63*, 1–11. [CrossRef] [PubMed]
24. Cavanagh, J.F. EEG: Probabilistic Selection and Depression. 2021. Available online: https://openneuro.org/datasets/ds003474/versions/1.1.0 (accessed on 26 August 2021). [CrossRef]
25. Nolan, H.; Whelan, R.; Reilly, R.B. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *J. Neurosci. Methods* **2010**, *192*, 152–162. [CrossRef] [PubMed]
26. Cavanagh, J.F.; Masters, S.E.; Bath, K.; Frank, M.J. Conflict acts as an implicit cost in reinforcement learning. *Nat. Commun.* **2014**, *5*, 5394. [CrossRef]
27. Broadway, J.M.; Frank, M.J.; Cavanagh, J.F. Dopamine D2 agonist affects visuospatial working memory distractor interference depending on individual differences in baseline working memory span. *Cogn. Affect. Behav. Neurosci.* **2018**, *18*, 509–520. [CrossRef]
28. Cavanagh, J.F.; Frank, M.J.; Broadway, J. EEG: Visual Working Memory + Cabergoline Challenge. *OpenNeuro* **2021**. [CrossRef]
29. Pavlov, Y.G. EEG: verbal working memory. *OpenNeuro* **2021**. [CrossRef]
30. Palmer, J.; Kreutz-Delgado, K.; Makeig, S. *AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components*; Technical Report; Swartz Center for Computational Neuroscience: San Diego, CA, USA, 2011; pp. 1–15.
31. Pavlov, Y.G.; Kotchoubey, B.; Pavlov, Y.G. Temporally distinct oscillatory codes of retention and manipulation of verbal working memory Corresponding author. *bioRxiv* **2021**. [CrossRef]
32. Tadel, F.; Baillet, S.; Mosher, J.C.; Pantazis, D.; Leahy, R.M. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Comput. Intell. Neurosci.* **2011**, *2011*, 879716. [CrossRef]
33. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **2010**, *52*, 1059–1069. [CrossRef]
34. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
35. Gao, S.; Calhoun, V.D.; Sui, J. Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci. Ther.* **2018**, *24*, 1037–1052. [CrossRef]
36. Dhanapal, R.; Bhanu, D. Electroencephalogram classification using various artificial neural networks. *J. Crit. Rev.* **2020**, *7*, 891–894. [CrossRef]
37. Jiao, Z.; Gao, X.; Wang, Y.; Li, J.; Xu, H. Deep Convolutional Neural Networks for mental load classification based on EEG data. *Pattern Recognit.* **2018**, *76*, 582–595. [CrossRef]
38. Kuanar, S.; Athitsos, V.; Pradhan, N.; Mishra, A.; Rao, K.R. Cognitive Analysis of Working Memory Load from Eeg, by a Deep Recurrent Neural Network. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings, Calgary, AB, Canada, 15–20 April 2018; pp. 2576–2580. [CrossRef]
39. Bilucaglia, M.; Duma, G.M.; Mento, G.; Semenzato, L.; Tressoldi, P. Applying machine learning EEG signal classification to emotion-related brain anticipatory activity. *F1000Research* **2020**, *9*, 173. [CrossRef]
40. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [CrossRef]
41. Medvedev, A.V.; Agoureeva, G.I.; Murro, A.M. A Long Short-Term Memory neural network for the detection of epileptiform spikes and high frequency oscillations. *Sci. Rep.* **2019**, *9*, 19374. [CrossRef]
42. Pavlov, Y.G.; Kotchoubey, B. The electrophysiological underpinnings of variation in verbal working memory capacity. *Sci. Rep.* **2020**, *10*, 16090. [CrossRef]

43. Patel, M.J.; Khalaf, A.; Aizenstein, H.J. Studying depression using imaging and machine learning methods. *Neuroimage Clin.* **2016**, *10*, 115–123. [CrossRef] [PubMed]
44. Pessoa, L. Understanding emotion with brain networks. *Curr. Opin. Behav. Sci.* **2018**, *176*, 19–25. [CrossRef] [PubMed]
45. Cavanagh, J.F.; Bismark, A.W.; Frank, M.J.; Allen, J.J.B. Multiple Dissociations Between Comorbid Depression and Anxiety on Reward and Punishment Processing: Evidence From Computationally Informed EEG. *Comput. Psychiatry* **2019**, *3*, 1. [CrossRef] [PubMed]
46. Zaharchuk, H.A.; Karuza, E.A. Multilayer networks: An untapped tool for understanding bilingual neurocognition. *Brain Lang.* **2021**, *220*, 104977. [CrossRef] [PubMed]
47. Sharma, A.; Verbeke, W.J.M.I. Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Front. Big Data* **2020**, *3*, 15. [CrossRef] [PubMed]

big data and cognitive computing

MDPI

Article

Exploring How Phonotactic Knowledge Can Be Represented in Cognitive Networks

Michael S. Vitevitch *, Leo Niehorster-Cook and Sasha Niehorster-Cook

Department of Psychology, University of Kansas, Lawrence, KS 66045, USA; leoniehorstercook@gmail.com (L.N.-C.); sniehorstercook@gmail.com (S.N.-C.)
* Correspondence: mvitevit@ku.edu

Abstract: In Linguistics and Psycholinguistics, phonotactics refers to the constraints on individual sounds in a given language that restrict how those sounds can be ordered to form words in that language. Previous empirical work in Psycholinguistics demonstrated that phonotactic knowledge influenced how quickly and accurately listeners retrieved words from that part of memory known as the mental lexicon. In the present study, we used three computer simulations to explore how three different cognitive network architectures could account for the previously observed effects of phonotactics on processing. The results of Simulation 1 showed that some—but not all—effects of phonotactics could be accounted for in a network where nodes represent words and edges connect words that are phonologically related to each other. In Simulation 2, a different network architecture was used to again account for some—but not all—effects of phonotactics and phonological neighborhood density. A bipartite network was used in Simulation 3 to account for many of the previously observed effects of phonotactic knowledge on spoken word recognition. The value of using computer simulations to explore different network architectures is discussed.

Keywords: phonotactic probability; neighborhood density; sub-lexical representations; lexical representations; phonemes; biphones; network science; cognitive network

Citation: Vitevitch, M.S.; Niehorster-Cook, L.; Niehorster-Cook, S. Exploring How Phonotactic Knowledge Can Be Represented in Cognitive Networks. *Big Data Cogn. Comput.* **2021**, *5*, 47. https://doi.org/10.3390/bdcc5040047

Academic Editors: Massimo Stella and Yoed N. Kenett

Received: 23 June 2021
Accepted: 15 September 2021
Published: 23 September 2021

1. Introduction

In Linguistics and Psycholinguistics, phonotactics refers to the individual sounds (known as phonemes) that are used in a given language, as well as the constraints on how those sounds can be ordered to form words in that language [1–3]. For example, Arabic uses a glottal stop /ʔ/, but English does not. (Characters or sequences of characters placed between angled lines (i.e., //) are symbols from the International Phonetic Alphabet (IPA), which are used to represent the sounds found in the languages of the world.) Similarly, the sequence /br/ is permissible at the ends of words in Arabic, but is not permissible at the beginning of words. The opposite is true for the sequence /br/ in English.

Speakers of a given language are not only aware of which phonemes and sequences of phonemes are used in the language(s) they speak, but they are also implicitly aware that certain phonemes and sequences are more common than other phonemes and sequences in the language. The variability in the frequency with which phonemes and sequences occur in a language is referred to as phonotactic probability. For example, in English words the phoneme /p/ and the sequence /pæv/ ("pav") occurs often in words, and would be said to have high phonotactic probability. In contrast, the phoneme /ʒ/ and the sequence /ðeʒ/ ("thayzh") occurs less often in English words and would be said to have low phonotactic probability.

Research in Linguistics and Psycholinguistics has shown that by 9-months of age children prefer to listen to nonwords containing high—rather than low—probability phonemes and sequences of phonemes, demonstrating that sensitivity to the sounds of one's native language occurs early in life [4]. Additional research has found that listeners rely on phonotactic probability to segment individual words from the stream of fluent speech [5],

to recognize words in speech [1,2], and to learn new words [6]. For a review of research on how phonotactic probability influences spoken word recognition and other language processes see [7].

Seminal work was conducted by [1,2] on how phonotactic probability influences spoken word recognition. They found that when participants were asked to repeat the words and nonwords that they heard (in a psycholinguistic task known as an auditory naming task), participants responded differentially as a function of the phonotactic probability of the words that naturally varied in phonotactic probability and the nonwords that were specially created to vary in phonotactic probability (such as the examples /pæv/ and /ðeʒ/ described above). As shown in Figure 1, participants responded more quickly and accurately to words with low phonotactic probability than to words with high phonotactic probability. For the nonwords, however, participants responded more quickly and accurately to nonwords with high phonotactic probability than to nonwords with low phonotactic probability.

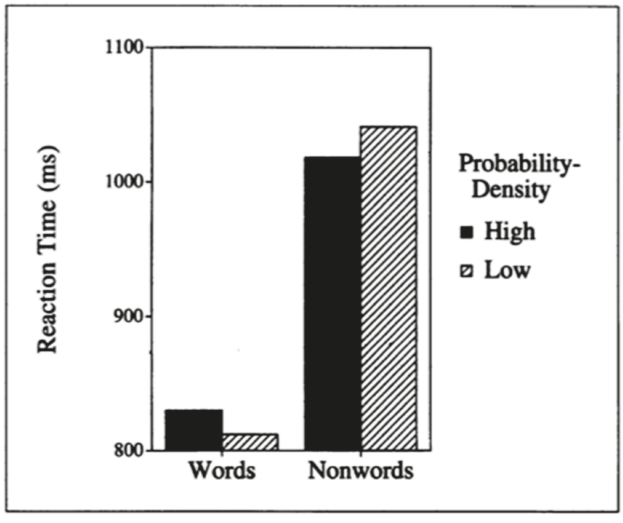


Figure 1. The response times in milliseconds of participants responding to words and nonwords in an auditory naming task. "Probability-density" refers to the terms phonotactic probability and neighborhood density that are used in the text. Used with permission from [1].

These results were interpreted by [1,2] as indicative that listeners represented in the mental lexicon (i.e., that part of memory that stores linguistic knowledge) not only information about the words they knew, but also information about the phonemes and sequences of phonemes that occur in the language. When the representations of phonemes and sequences of phonemes (so called, "sub-lexical representations") were used to process the spoken input, one could expect to see stimuli with high phonotactic probability being responded to more quickly and accurately than stimuli with low phonotactic probability. Enhanced performance to stimuli that are common in the environment are common in many domains of perceptual and cognitive processing [8], and may give rise to the processing advantage observed for stimuli with high phonotactic probability.

However, when representations of words (so called, "lexical representations") were used to process the spoken input, one could expect to see stimuli with low phonotactic probability being responded to more quickly and accurately than stimuli with high phonotactic probability. This is because words with high phonotactic probability, like *cat*, are confusable with many words in the language (e.g., *rat*, *fat*, *mat*, *sat*, *hat*, *cut*, *kit*, *cap*, *can*, *calf*, etc.), and words with low phonotactic probability, like *dog*, are confusable with fewer

words in the language (e.g., *log*, *hog*, *dig*, *doll*). A word that is confusable with many words (e.g., *cat*) is said to have high neighborhood density, whereas a word that is confusable with fewer words (e.g., *dog*) is said to have low neighborhood density. Numerous studies have demonstrated that when there are many confusable words to discriminate among, spoken word recognition occurs slowly and less accurately (for reviews see [9–11]).

Participants can be induced to use lexical representations by presenting them with real words in a psycholinguistic task, or by asking them to engage in a task, such as a lexical decision task, that requires them to discriminate whether a stimulus is a real word or a nonword. In tasks that induce the listener to use lexical representations, one would expect to find that words with high neighborhood density (e.g., *cat*) are responded to slower and less accurately than words with low neighborhood density (e.g., *dog*).

Participants can be induced to use sub-lexical representations by presenting them with nonwords and by asking them to engage in a task, such as the auditory naming task, that does not require them to discriminate among words (i.e., they simply need to repeat what they hear, whether it is a word or not). Participants can also be induced to use sub-lexical representations to process real words when a small number of real words are embedded with many nonwords in a task that does not require them to discriminate among words, such as a same-different decision of two sequentially presented stimuli [12]. In tasks that induce the listener to use sub-lexical representations, words may be treated as if they are nonwords, so we would expect to find that nonwords and words with high phonotactic probability are responded to more quickly and accurately than nonwords and words with low phonotactic probability.

To account for these effects, [2] appealed to a type of artificial neural network called adaptive resonance theory (ART; [13]), because it included sub-lexical representations that were sensitive to how common they occurred in the environment (i.e., producing enhanced performance for more common items), and lexical representations that were sensitive to how many other similar sounding words they could be confused with (i.e., a word that is confusable with few other words is responded to quickly and accurately). Subsequent computer simulations by [14] formally confirmed that an ART network could account for the results observed by [1,2]. Note that the ART network and other types of artificial neural networks (e.g., Kohonen neural networks, recurrent neural networks, etc.) differ from the cognitive networks used in the present simulations. In the present simulations, we explored if a different type of network—namely, cognitive networks—could represent phonotactic information in some way to also account for the results observed by [1,2]. The term "cognitive network" has emerged recently [15] to describe applications of the mathematical tools of network science to questions commonly studied by cognitive psychologists and cognitive scientists [15,16]). Broadly speaking, artificial neural networks such as ART attempt to model cognitive processing, whereas cognitive networks attempt to capture how representations are organized in memory. It is important in the cognitive network approach to understand how representations are organized in memory because that structure can make cognitive processes more or less efficient.

An example of a cognitive network is illustrated in Figure 2, where nodes represent words in the mental lexicon, and edges connect words that are phonologically related [17]. Cognitive networks can also be constructed of words that are semantically related [18], and, as in the simulations reported below, of sub-lexical representations instead of whole words.

In what follows we report the results of several computer simulations that explored how three different cognitive network architectures might account for the effects of phonotactic probability and neighborhood density on words and nonwords that were observed by [1,2]. The advantages and disadvantages of each cognitive network architecture are also discussed.

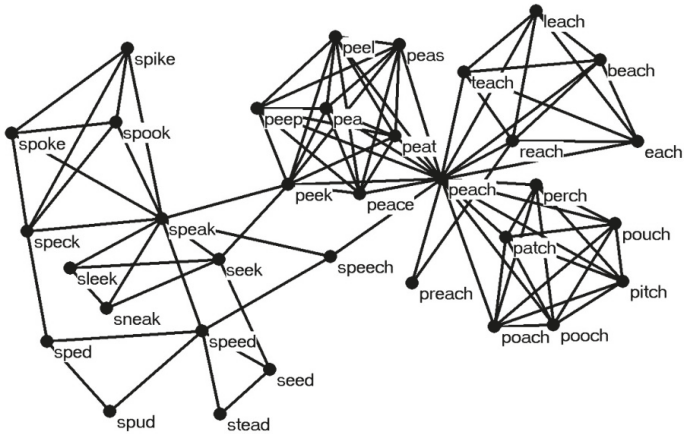


Figure 2. An example of a cognitive network in which nodes represent words in the mental lexicon, and edges connect words that are phonologically similar to each other (based on the addition, deletion, or substitution of a phoneme in one word to form another word). Phonological similarity can be defined in other ways as well.

2. Simulation 1: Lexical Network

In the first simulation, we used the phonological network first examined by [17]. In that network, nodes represent 19,340 words in the mental lexicon, and edges connect words if the addition, deletion, or substitution of a phoneme in one word formed the other word (as in Figure 2). For example, deleting the /s/ in the word *speech* produces the phonologically related word *peach*.

Note that there are no sublexical representations—neither individual phonemes nor biphones—explicitly represented in this network. We were motivated to examine the extent to which a cognitive network containing only words could account for the effects of phonotactic probability and neighborhood density by two interesting findings.

First, [19] showed that phonotactic knowledge could emerge from lexical representations in their TRACE model of spoken word recognition. In this artificial neural network, it was found that phonotactic knowledge emerged as a result of a conspiracy effect among lexical representations. That is, the TRACE model would "know" that /tl-/ was not a legal sequence of word-initial phonemes in English, but that /tr-/ was a legal sequence of word-initial phonemes in English simply because there were several lexical representations that started with /tr–/, but none that started with /tl–/. The model did not explicitly represent information related to the legality or probability of sequences of phonemes, but such knowledge could emerge from the lexical representations.

Second, [20] used the Louvain method with a resolution of 0.2 to examine the community structure of the cognitive network of words examined by [17]. A modularity value, *Q*, of 0.655 was found, indicating a reasonably robust community structure. Further analysis of the communities, or sub-groups of words that were more likely to be connected to each other than to other words in the network, revealed that the words in a community tended to share common phonemes and sequences of phonemes. For example, one of the communities of words observed by [20] contained the biphones /■k/, /I■/ and /rI/, which were found in the words *rink*, *bring*, *drink* and *wrinkle* that populated that community. It was suggested that if cognitive processing focused on the individual word node, one would observe the neighborhood density effects typically observed in other studies (e.g., [9]). However, if cognitive processing was distributed across multiple word nodes, perhaps considering the overall activation of a community, one might observe the sub-lexical effects reported in [1,2]. In the present simulation we used an R package called *spreadr* [21] to diffuse activation across the lexical network of [17] for several time-steps. Although there

are other R packages that implement diffusion mechanisms in networks (e.g., [22]), *spreadr* implements the diffusion of activation in a way that is consistent with how it is commonly discussed and used in the cognitive sciences. In this context, activation is viewed as a limited cognitive resource that can spread to and activate the information in connected nodes (e.g., [23]).

The real words varying in neighborhood density and phonotactic probability that were used as stimuli in [2] were presented to the network model. To examine cognitive processing of the individual word, we examined the activation level of those stimulus words at the end of five time-steps during which activation had diffused across the network. To examine cognitive processing that was distributed across multiple word nodes, we considered the sum of the activation levels of all of the other words in the network that had been partially activated at the end of five time-steps. Note that the sum of the activation of all the words that had been activated after five time-steps results in processing that is more distributed than simply considering the activation levels of the other words in a given community. We elected to use all of the partially activated words rather than just the words in the community to which the stimulus word belonged in order to strike a balance between the idea from [20] of distributed processing, and the idea from [19] of conspiracy effects emerging from all of the words in the lexicon.

If [20] is correct, then when we examine the activation of the individual stimulus words, we should find that words that are low in neighborhood density/phonotactic probability will have higher activation levels (corresponding to faster and more accurate performance in humans) than words that are high in neighborhood density/phonotactic probability. Recall that words with high phonotactic probability tend to have high neighborhood density (and words with low phonotactic probability tend to have low neighborhood density). Henceforth, we will use the terms neighborhood density/phonotactic probability to emphasize effects driven by lexical representations, and phonotactic probability/neighborhood density to emphasize effects driven by sub-lexical representations. This result would replicate the effects observed by [1,2] for words that had been processed with lexical representations.

Further, when we examine the sum of the activation levels of all of the other words in the network that had been partially activated at the end of five time-steps, we should now find that words that are high in phonotactic probability/neighborhood density will have higher activation levels (corresponding to faster and more accurate performance in humans) than words that are low in phonotactic probability/neighborhood density. This result would replicate the effects observed by [12] for words that had been processed using sub-lexical representations.

2.1. Materials and Methods of Simulation 1

The network used in the present simulation consisted of the 19,340 words in the phonological network examined in [17]. Edges were placed between word nodes if the addition, deletion, or substitution of a single phoneme in one word resulted in the other word. The 140 real words from [2] were presented to *spreadr* ([21]; version 0.2.0) with the following settings for the various parameters in the model. An initial activation value of 20 units was used for each stimulus word in the present simulation. Our decision to use an initial activation value of 20 is arbitrary, and qualitatively similar results would be obtained using other initial activation values (e.g., [24]).

Decay (d) refers to the proportion of activation lost at each time step. This parameter ranges from 0 to 1, and was set to 0 in the simulations reported here to be consistent with the parameter settings used in previous simulations (e.g., [24,25]).

Retention (r) refers to the proportion of activation retained in a given node when it diffused activation evenly to the other nodes connected to it. This value ranges from 0 to 1, and was set to 0.5 in the simulations reported here. In [24] values ranged from 0.1 to 0.9 in increments of 0.1. Because the various retention values in [24] produced comparable results across retention values, we selected in the present simulation a single, mid-range value

(0.5) for the retention parameter in order to reduce the computational burden, thereby accelerating data collection.

The *suppress* (*s*) parameter in *spreadr* forces nodes with activation values lower than a selected value to *activation* = 0. It was suggested in [21] that when this parameter is used a very small value (e.g., <0.001) should be used. In the present simulations *suppress* = 0 in order to be consistent with the parameter settings used in previous simulations (e.g., [24,25]).

Time (*t*) refers to the number of time steps that activation diffuses or spreads across the network. In the present simulations $t = 5$. This value was selected because as shown in Figure 3 of [21], activation values reach asymptote in approximately five time-steps. Furthermore, as shown in the hop-plot depicted in Figure 2 of [26] approximately 50% of the network has been reached by traversing on average five connections (i.e., hops) in every direction from a given node, suggesting that the network has been sufficiently saturated. This value for the time parameter ($t = 5$) enabled us to reduce the computational burden, thereby accelerating data collection. At the end of five timesteps we documented the activation level of each of the stimulus words, and summed the activation of all the other words that had been partially activated.

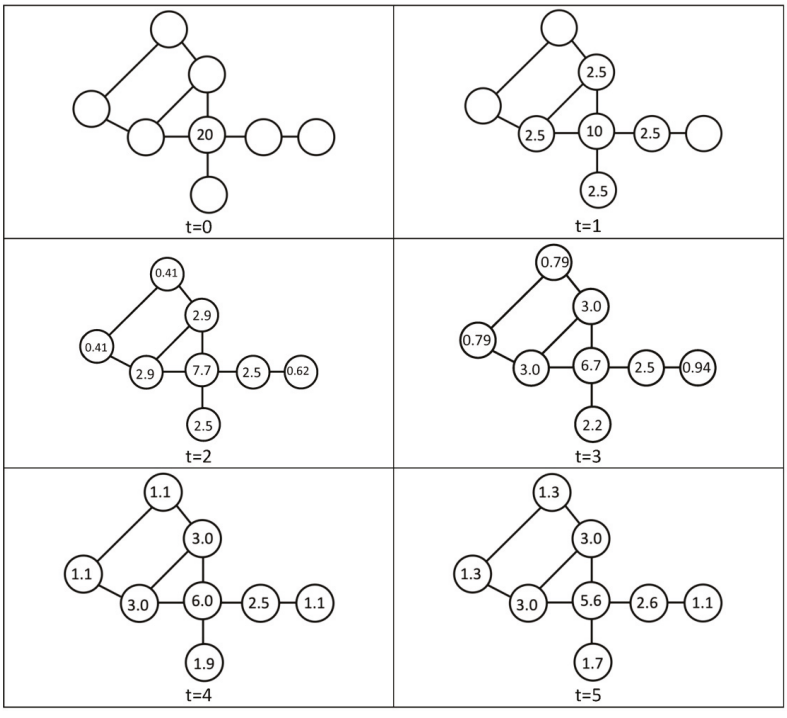


Figure 3. A simplified example of how activation diffuses for 5 time-steps across the network using the same parameters as used in the simulations reported here. $t = 0$ refers to the initial activation of the stimulus node with 20 units of activation. Nodes with no values have activation = 0. Note that activation values may have changed at decimal positions not shown in the figure.

Details about the words and nonwords used in the present simulations can be found in [1,2], but we provide here a few key characteristics of the stimuli. The 140 real words and 240 nonwords were monosyllabic, both had a consonant–vowel–consonant syllable structure. Phonotactic probability was calculated as in [27,28] taking in to account the frequency of occurrence of the segments and the biphones in the words and the nonwords. A median split was used to categorize the words and nonwords into the high and low phonotactic

probability categories used in [1,2] and in the present simulations. A similar procedure (i.e., median split) was used to categorize the words and nonwords in to high and low neighborhood density categories. The words in the high categories were matched to words in the low categories on a number of factors including the initial phonemes of the words, stimulus duration, and frequency of occurrence in English. The nonwords in the high categories were matched to the nonwords in the low categories in a comparable manner.

2.2. Results of Simulation 1

In [1,2] words with low neighborhood density/phonotactic probability like *dog* were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability like *cat*. In the cognitive network model implemented on *spreadr*, we found when looking at the activation values of the stimulus words that words with low neighborhood density/phonotactic probability had higher activation levels (*mean* = 1.23 units; *sd* = 0.31) indicating that they were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability (*mean* = 1.14 units; *sd* = 0.14). Similar to the analyses used in [1,2], an independent samples *t*-test, a statistic that is robust to various assumption violations [29], was used [30] and indicated that this difference was statistically significant (t (138) = −2.29, $p < 0.05$). This result qualitatively replicates the results of [1,2].

As described above participants in [12] were induced to respond to words using sub-lexical representations, resulting in words with high phonotactic probability/neighborhood density being responded to more quickly and accurately than words with low phonotactic probability/neighborhood density. In the cognitive network model implemented on *spreadr*, we found when looking at the sum of the activation values of all of the other words in the network that had been partially activated at the end of five time-steps that words in the high phonotactic probability/neighborhood density condition had higher activation levels (*mean* = 18.86 units; *sd* = 0.14) indicating they were responded to more quickly and accurately than words with low phonotactic probability/neighborhood density (*mean* = 18.77 units; *sd* = 0.31). An independent samples *t*-test indicated that this difference was statistically significant (t (138) = 2.29, $p < 0.05$). This result qualitatively replicates the results of [12].

2.3. Discussion of Simulation 1

Based on the ideas of [19,20], see also [31], that phonotactic knowledge can emerge from lexical representations, we examined if a cognitive network that contained only words might be able to exhibit knowledge of phonotactic information. Previous attempts to account for the representation of phonotactic information have typically appealed to the explicit representation of words (i.e., lexical representations) and phonemes and biphones (i.e., sub-lexical representations). To examine if phonotactic knowledge could emerge from a cognitive network of only words we simulated the retrieval of the words from [1,2] in the phonological network of [17].

We predicted that when we examined the activation of the individual stimulus words, we should find that words that are low in neighborhood density/phonotactic probability will have higher activation levels (corresponding to faster and more accurate performance in humans) than words that are high in neighborhood density/phonotactic probability as observed in [1,2]. We further predicted that when we examined the sum of the activation levels of all of the other words in the network that had been partially activated at the end of five time-steps, we should find that words in the high phonotactic probability/neighborhood density condition will have higher activation levels (corresponding to faster and more accurate performance in humans) than words in the low phonotactic probability/neighborhood density condition, as observed by [12] for words that had been processed using sub-lexical representations. The results of Simulation 1 confirmed those predictions and provide some credence to the idea that phonotactic knowledge can emerge from lexical representations.

The results of Simulation 1 also demonstrate that some forms of phonotactic knowledge can be represented using the cognitive network science approach. This is an important demonstration because in order for the cognitive network approach to be a useful approach it should be able to account for a wide range of linguistic phenomena. The results of Simulation 1, therefore, represent an important step in that direction.

It is also important to acknowledge the limitations of the present simulation. Although lexical representations are typically used during spoken word recognition [1,2], there are other language processes that rely on sub-lexical representations, such as segmenting individual words from the stream of fluent speech [5], and learning new words [6]. Given that there are only words in the cognitive network used in the present simulation, there is no way to examine how these other linguistic phenomena might be accounted for by the cognitive network approach with the present network architecture. Indeed, there is also no way to examine the processing of the nonwords used by [1,2], because, by definition, nonwords are not represented in the lexicon. Therefore, in the simulations that follow, we explored different cognitive network architectures that included biphones (Simulation 2) and phonemes (Simulation 3).

3. Simulation 2: Network of Words Connected via Shared Biphones

Recall that the network used in Simulation 1 contained words that were connected based on the addition, deletion, or substitution of a single phoneme [17]. For example, the words *cat* (/kæt/) and *cut* (/kʌt/) were connected because substitution of the medial phoneme resulted in the other word. Even though the network did not have explicit knowledge of phonemes and biphones, the results of Simulation 1 showed that phonotactic knowledge could emerge from the lexicon. In the present simulation, we used a different network architecture to explicitly include knowledge of biphones in the cognitive network.

In the network used in the present simulation, nodes again represented the 19,340 words from the network used in Simulation 1. In this case, however, edges connected words if they contained one or more shared biphones (see [32] for networks constructed using a similar approach). As shown in Figure 4, for example, the word *clean* (/klin/) contains the biphones /kl/, /li/, and /in/. Edges in the present network would connect *clean* to all the other words that contain at least one of those biphones, such as *flea* (/fli/), *teens* (/tinz/), *forklift* (/fɔklɪft/) and *ballerina* (/bələɹinə/). The word *clean* (/klin/) would not be connected to words that merely contain some of the same phonemes, such as *link* (/lInk/) or *alcohol* (/ælkəhɑl/), both of which contain /k/ and /l/, but not the specific biphone /kl/. Although the words *cat* (/kæt/) and *cut* (/kʌt/) were connected in the network used in Simulation 1 [17], these words share no biphones, and are therefore not connected in the present network.

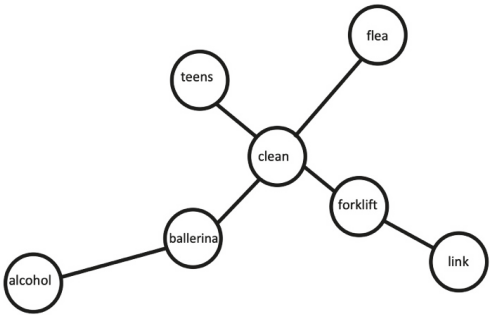


Figure 4. An example of several words in the network used in Simulation 2 in which nodes represent words in the mental lexicon, and edges connect words that are phonologically similar to each other (based on sharing a biphone).

As in Simulation 1, the *R* package *spreadr* [21] was again used to diffuse activation across the network (using the same parameters as in Simulation 1). This allowed us to examine how a cognitive network with a different architecture might account for the effects of phonotactic probability and neighborhood density on the words originally tested in psycholinguistic experiments by [1,2].

3.1. Materials and Methods for Simulation 2

The network used in the present simulation contained the same 19,340 words used in Simulation 1 [17]. In Simulation 1, words were connected based on the addition, deletion or substitution of a single phoneme. In the present case, words were connected if they had one or more biphones in common. The networks used in the present simulation and in Simulation 3 were generated in R version 4.1.0 [33] using the following libraries: igraph version 1.2.6 [34], and tidyverse [35].

Although the present simulation used a new method by which to connect words, the degree (which corresponds to neighborhood density) of the words in the present network is qualitatively similar to the degree of the words in the network used in Simulation 1. For the network in Simulation 1, the words with high neighborhood density/phonotactic probability had a *mean* degree of 26.9 words (*sd* = 5.9), and the words with low neighborhood density/phonotactic probability had a *mean* degree of 20.0 words (*sd* = 7.0). For the network in Simulation 2 the words with high neighborhood density/phonotactic probability had a *mean* degree of 565.8 words (*sd* = 329.7), and the words with low neighborhood density/phonotactic probability had a *mean* degree of 258.6 words (*sd* = 182.9). A Pearson's correlation shows that the measure of degree from the network used in Simulation 1 is correlated with the measure of degree from the network used in Simulation 2 (r (138) = 0.47, $p < 0.001$), suggesting that the two ways of placing edges between words in the two networks are similar. The *R* package *spreadr* [21] was again used to diffuse activation across the network. The same parameters used in Simulation 1 (*decay* = 0, *retention* = 0.5, *suppress* = 0, $t = 5$) were also used in the present simulation.

As in Simulation 1, the 140 real words varying in neighborhood density/phonotactic probability from [2] were again presented to the network. We again examined the activation value of each word at the end of five time-steps.

3.2. Results of Simulation 2

In [1,2] words with low neighborhood density/phonotactic probability were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability. In the cognitive network model implemented on *spreadr* with words connected because they shared biphones, we found when looking at the activation values of the stimulus words that words with low neighborhood density/phonotactic probability had higher activation levels (*mean* = 0.638 units; *sd* = 0.009) indicating that they were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability (*mean* = 0.632 units; *sd* = 0.003). An independent samples t-test indicated that this difference was statistically significant (t (138) = −5.32, $p < 0.001$). This result qualitatively replicates the results of [1,2] and of Simulation 1 using a different way to represent phonotactic knowledge in the cognitive network architecture.

As in Simulation 1, we sought in the present simulation to examine the ideas of [19,20,31] that phonotactic knowledge can emerge from lexical representations. We again examined if the sum of the other activated words in the network would capture how words are processed if sublexical representations were used.

We found when looking at the sum of the activation values of all of the other words in the network that had been partially activated at the end of five time-steps that words in the high phonotactic probability/neighborhood density condition had higher activation levels (*mean* = 19.368 units; *sd* = 0.003) indicating they were responded to more quickly and accurately than words with low phonotactic probability/neighborhood density (*mean* = 19.362 units; *sd* = 0.009). An independent samples t-test indicated that this

difference was statistically significant (t (138) = 5.32, $p < 0.001$), which qualitatively replicates the results of [12]. The different architecture for the cognitive network used in the present simulation also allows us to assess how the number of phonological neighbors affects processing using a different definition of phonological similarity. Recall that the number of phonologically related words is often determined by counting the number of words that are similar to a given word based on the addition, deletion, or substitution of a phoneme in one word to form another word [9,17]. However, phonological similarity has been defined in other ways as well [9,10].

In the present simulation, we can assess how the number of phonological neighbors influences processing using a different measure of phonological neighbor, namely the degree of each word, which in the present network indicates that two words share one or more biphones. A Pearson's correlation shows a significant relationship between the number of neighbors a word has (i.e., degree based on sharing at least one biphone) and the activation value of the word after five time-steps (r (138) = −0.566, $p < 0.001$). That is, words with few neighbors had higher activation values indicating they were responded to more quickly and accurately than words with many neighbors. This result replicates previous findings about the influence of the number of neighbors on spoken word recognition that used different ways to define phonological similarity [9,10].

3.3. Discussion of Simulation 2

In the present simulation we used a different architecture to represent phonotactic information in the cognitive network. Specifically, word nodes were connected if they shared one or more biphones, thereby explicitly encoding phonotactic knowledge in the cognitive network in the edges between word nodes. This representational scheme differs from the architecture used in Simulation 1, where a word node was connected to another word node if the addition, deletion, or substitution of a single phoneme produced the other word. Despite using a different network architecture in the present simulation, we were still able to qualitatively replicate the results of [1,2]. That is, real English words with low neighborhood density/phonotactic probability were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability. Further, when, as in Simulation 1, we summed the activation of the other partially activated words to assess the emergence of phonotactic effects from lexical representations we found that real English words with high phonotactic probability/neighborhood density were responded to more quickly and accurately than words with low phonotactic probability/neighborhood density, qualitatively replicating the results of [12].This new representational scheme also provided a different way to define phonological similarity among words, one that differs from the one-phoneme metric often used to define the phonological neighborhood [9]. Despite phonological similarity now being defined by the sharing of biphones, we also replicated the influence of phonological neighborhood density in spoken word recognition. Specifically, words with few phonological neighbors (i.e., low neighborhood density, or low degree in the network) are responded to more quickly and accurately than words with many phonological neighbors (i.e., high neighborhood density, or high degree in the network).

Although phonotactic knowledge was now explicitly encoded in the cognitive network via the edges between words, we, like in Simulation 1, were still not able to test how the cognitive network performs on the nonword stimuli used in [1,2]. The influence of sub-lexical representations such as phonemes and biphones appears limited in spoken word recognition [1,2]. However, there are other language processes that do rely on sub-lexical representations, such as segmenting individual words from the stream of fluent speech [5], and learning new words [6]. Further, [1,2] observed differences in performance in response to the nonwords that they constructed to vary in phonotactic probability. To more fully examine if the cognitive network approach can account for the influence of phonotactic knowledge on spoken word recognition and perhaps other language processes, we used yet another network architecture in Simulation 3. This new architecture will allow us to

attempt to replicate via computer simulation the results observed by [1,2] for the real word as well as the nonword stimuli.

4. Simulation 3: Network Containing Phonemes and Words

In Simulations 1 and 2 the cognitive network contained only lexical nodes (i.e., they were one mode networks). In Simulation 1, phonologically similar word nodes were connected based on a one-phoneme metric. Thus, phonotactic knowledge was not explicitly encoded in the network, but, as suggested by [19,20], phonotactic knowledge can emerge from subsets of words in the lexicon. In Simulation 2, phonologically similar word nodes were connected if they shared one or more biphones. In that representational scheme, which was a one mode projection of a bipartite network containing biphones and word nodes, some phonotactic knowledge was explicitly encoded in the edges between nodes in the network.

In the present simulation we wished to encode phonotactic knowledge by including both phoneme nodes and word nodes in the network. However, in contrast to the one-mode projection used in Simulation 2, in the present simulation we retained the bipartite structure. Thus, word nodes were not directly connected to each other, and phoneme nodes were not directly connected to each other. Rather phoneme nodes connected to word nodes if the word contained that phoneme.

Consider the word nodes for *cat* (/kæt/) and *cut* (/kʌt/) as shown in Figure 5. In the network architecture used in Simulation 1, those word nodes would be connected by an edge, because of a single phoneme substitution of the medial segment. In the network architecture used in Simulation 2, those word nodes would not be connected by an edge, because those two words do not share a common biphone. In the bipartite network used in the present simulation *cat* and *cut* would again not be directly connected by an edge. Instead, nodes for the phonemes /k/ and /t/ would connect to the *cat* and *cut* nodes. Only the node for the phoneme /æ/ would connect to the node for *cat*, and only the node for the phoneme /ʌ/ would connect to the node for *cut*.

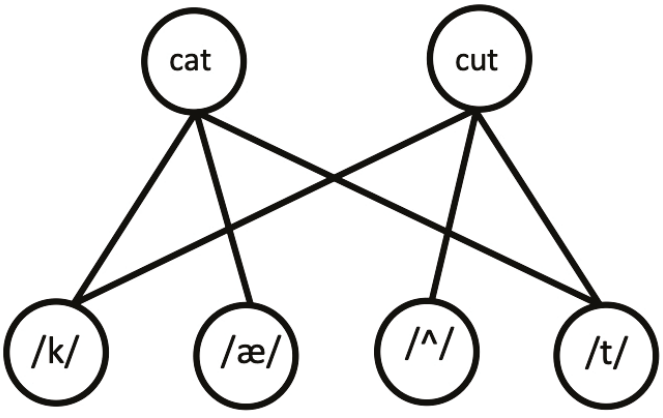


Figure 5. An example of two words and four phonemes in the bipartite network used in Simulation 3. Nodes represented words and phonemes, and edges connected phonemes to words that contained those phonemes. Words were not directly connected to each other as they were in the networks used in Simulations 1 and 2.

This bipartite network architecture allowed us to examine processing when lexical representations were used to respond to real words (as in Simulations 1 and 2) by considering the final activation levels of the word nodes. This network architecture also allowed us to examine processing when sublexical representations were used to respond to both real

words [12] and nonwords [1,2] by considering the sum of the final activation levels of the phoneme nodes.

4.1. Materials and Methods of Simulation 3

The word nodes in the network used in the present simulation were the same 19,340 words used in Simulations 1 and 2 [17]. The phoneme nodes in the present bipartite network were the 45 phonemes used to transcribe the words. Word nodes were not directly connected to each other, and phoneme nodes were not directly connected to each other. Rather, edges appeared between phoneme nodes and word nodes if the word contained those phonemes, thereby forming a bipartite network.

The *R* package *spreadr* [21] was again used to diffuse activation across the network. The same parameters used in Simulations 1 and 2 were also used in the present simulation, except that activation was allowed to spread over twice as many time-steps given the increased distance between lexical representations resulting from the bipartite architecture of the network (*decay* = 0, *retention* = 0.5, *suppress* = 0, $t = 10$). To examine cognitive processing that uses lexical representations, activation started at the word node corresponding to the target word, and diffused across the network. After 10 time-steps, activation values in the nodes corresponding to the target words were then recorded and analyzed. To examine cognitive processing that uses sublexical representations, activation started at the phoneme nodes that were constituents of the target word or target nonword, and diffused across the network. After 10 time-steps, activation values in the constituent phoneme nodes were then recorded and the sum of those values were analyzed.

As in Simulations 1 and 2, the 140 real words varying in phonotactic probability and neighborhood density from [2] were presented to the network. This time, however, we also used the 240 nonwords (120 were designated high phonotactic probability/neighborhood density, and 120 were designated low phonotactic probability/neighborhood density) from [1,2].

4.2. Results of Simulation 3

In [1,2], when lexical representations were used to process the real English words, words with low neighborhood density/phonotactic probability were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability. In the bipartite network model containing nodes for phonemes and words, we examined the activation values after 10 time-steps at the word nodes to simulate the use of lexical representations to process the words.

We found when looking at the activation values of the stimulus words that words with low neighborhood density/phonotactic probability had higher activation levels (*mean* = 0.021 units; *sd* = 0.0005086) indicating they were responded to more quickly and accurately than words with high neighborhood density/phonotactic probability (*mean* = 0.020 units; *sd* = 0.0002767). An independent samples *t*-test indicated that this difference was statistically significant (t (138) = −5.43, $p < 0.001$). This result qualitatively replicates the results of [1,2] and of Simulations 1 and 2 using a different way of representing phonotactic knowledge in the cognitive network.

In [12], when sublexical representations were used to process the real English words, words with high phonotactic probability/neighborhood density were now responded to more quickly and accurately than words with low phonotactic probability/neighborhood density, similar to the pattern typically seen for nonwords. In the bipartite network model containing nodes for phonemes and words, we examined the sum of the activation values after 10 time-steps at the phoneme nodes to simulate the use of sublexical representations to process the words.

We found when looking at the sum of the activation values of the phoneme nodes corresponding to the stimulus word that words with high phonotactic probability/neighborhood density had higher activation levels (*mean* = 4.190 units; *sd* = 0.682) indicating they were responded to more quickly and accurately than words with low phonotactic probabil-

ity/neighborhood density (*mean* = 3.516 units; *sd* = 0.727). An independent samples *t*-test indicated that this difference was statistically significant (t (138) = 5.66, $p < 0.001$). This result qualitatively replicates the results of [12].

We now consider the results for the nonwords. In [1,2], when sublexical representations were used to process the nonwords, nonwords with high phonotactic probability/neighborhood density were responded to more quickly and accurately than words with low phonotactic probability/neighborhood density. In the bipartite network model containing nodes for phonemes and words, we examined the sum of the activation values after 10 time-steps at the phoneme nodes to simulate the use of sublexical representations to process the nonwords.

We found when looking at the sum of the activation values of the phoneme nodes that were found in the nonword that nonwords with high phonotactic probability/neighborhood density had higher activation levels (*mean* = 3.708 units; *sd* = 0.607) indicating they were responded to more quickly and accurately than nonwords with low phonotactic probability/neighborhood density (*mean* = 2.150 units; *sd* = 0.372). An independent samples *t*-test indicated that this difference was statistically significant (t (238) = 23.97, $p < 0.001$). This result qualitatively replicates the results of [1,2].

4.3. Discussion of Simulation 3

In the present simulation we used a bipartite network containing nodes for phonemes and words. Phoneme nodes were not connected to each other, and word nodes were not connected to each other (in contrast to the word nodes in the networks used in Simulations 1 and 2). Rather, phoneme nodes connected to word nodes if that phoneme was a constituent of that word. This network architecture enabled us to simulate the use of lexical representations to process real words by considering the activation values at the word nodes after 10 time-steps.

Rather than looking at the emergence of phonotactic effects from a conspiracy of lexical representations as we did in Simulations 1 and 2, we were able in the present simulation using a different network architecture to examine the use of sublexical representations to process the real words and nonwords by considering the sum of the activation values at the phoneme nodes after 10 time-steps. The results of the present simulation using this bipartite representational scheme replicated the results for the words and nonwords observed in [1,2], and for the real words observed in [12].

5. Conclusions

We reported the results of computer simulations using three different network architectures to explore how the cognitive network approach might be used to represent phonotactic knowledge. Previous attempts to account for the processing of phonotactic information relied on artificial neural networks with lexical and sub-lexical representations [14]. However, more recent accounts have suggested that phonotactic knowledge could emerge just from lexical representations [20]. Therefore, in Simulation 1 we tested whether phonotactic information could emerge from a cognitive network of phonologically related words that did not have phonotactic information explicitly encoded in it, and did not have a separate layer of sub-lexical representations.

The simulation of lexical retrieval from the network of phonologically related words qualitatively replicated the results for real words observed by [1,2]. That is, when specific lexical representations were used to process the input, words with low neighborhood density/phonotactic probability had higher activation levels (indicating faster and more accurate responses) than words with high neighborhood density/phonotactic probability.

To examine the emergence of phonotactic information from collections of partially activated words in the lexicon we analyzed the sum of the activation levels of all of the other words in the network that had been partially activated at the end of five time-steps. In that analysis we found that words in the high phonotactic probability/neighborhood density condition had higher activation levels (corresponding to faster and more accurate

performance in humans) than words in the low phonotactic probability/neighborhood density condition, as observed by [12] for words that had been processed using sub-lexical representations. The results of this simulation lend credence to the idea that phonotactic knowledge can emerge solely from lexical representations. However, this network architecture which lacked sub-lexical representations (and of course did not, by definition, include nodes for nonwords in the lexicon) did not allow us to examine the effects observed by [1,2] for nonwords varying in phonotactic probability, prompting us to explore different network architectures in Simulations 2 and 3.

In Simulation 2 we explicitly modeled phonotactic knowledge in the network by connecting word nodes if they shared one or more biphones. Using this network architecture, we again found that words with low neighborhood density/phonotactic probability had higher activation levels (indicating faster and more accurate responses) than words with high neighborhood density/phonotactic probability, as observed by [1,2].

This network architecture also provided us with an alternative way to define phonological similarity that differed from the one-phoneme metric used to define phonological similarity among words in the network that was used in Simulation 1. Using this alterative definition of phonological similarity, we found that words with few phonological neighbors (i.e., low neighborhood density, or low degree in the network) were responded to more quickly and accurately than words with many phonological neighbors (i.e., high neighborhood density, or high degree in the network), replicating the often-observed effects of neighborhood density in spoken word recognition [9,10].

Finally, in Simulation 3 we used a bipartite network containing nodes for words and nodes for phonemes in order to explicitly represent phonotactic information with sub-lexical representations. This network architecture also allowed us to test whether the cognitive network approach could also account for the results observed by [1,2] for specially constructed nonwords in their psycholinguistic experiments. The results of the simulation using the bipartite representational scheme replicated the results for the words and nonwords observed in [1,2], and for the real words observed in [12], demonstrating that with the right network architecture the cognitive network approach can be used to account for the previously observed effects of phonotactic probability and neighborhood density on spoken word recognition. Future simulations could explore how these network architectures might account for the influence of phonotactic probability and neighborhood density in other language processes. The present set of simulations illustrates the value of using computer simulations to explore how different cognitive network architectures might or might not account for previously studied phenomena. For example, Simulations 1 and 2 contained lexical representations, but not sublexical representations. The network architectures in Simulations 1 and 2 allowed us to explore lexical effects (i.e., neighborhood density), and the idea that phonotactic information may emerge from a conspiracy of lexical representations. However, without sublexical representations in the networks used in Simulations 1 and 2, we were not able to directly examine some of the phonotactic probability effects that had been previously observed, especially those effects obtained with nonword stimuli (which by definition are not represented in the lexicon). The present work therefore highlights how important the network architecture is for accounting for a wide range of psycholinguistic phenomena.

For other demonstrations of cognitive networks accounting for other phenomena related to language and memory see [21]. We believe it is important for new approaches, such as cognitive networks, to be able to account for previous findings in new ways, as well as generate new ideas for future research.

The present exploration of alternative network architectures suggests that future work using cognitive networks may need to consider network architectures that are more complex in order to examine other types of linguistic phenomena. Those more complex network architectures may incorporate multiple layers of representation, like the bipartite network used in Simulation 3, or may take the form of multiplex networks. In multiplex networks there may be a network of phonologically related words connected to another

network layer in which the words are connected by edges if they are semantically related. Such networks have been used to examine a number of language phenomena, including the word-finding difficulties of people with aphasia (e.g., [36]).

Another approach may be to employ *feature-rich networks* [37], which are networks that include additional information (e.g., temporal or probabilistic information) to complement the topological information inherent in the network itself. One type of feature-rich network is a *node-attributed network*, which adds categorical or numerical information to the nodes. In the case of cognitive networks representing the mental lexicon, one might include information regarding the part of speech, length of the word, frequency of occurrence, or age of acquisition of the word nodes. Work on community detection algorithms in node-attribute networks has been able to successfully detect important modules that could not be detected by topological or homophilic clustering criteria alone [38], highlighting the insight that network science approaches might provide to language researchers. In addition to feature-rich networks, it might be worth exploring networks that grow over time. Such networks have been used to explore word learning and changes in the lexicon with age (*e.g.*, [39]). Much research has examined how phonotactic information influences word learning (*e.g.*, [6,40]). Exploring how phonotactic knowledge is represented in various types of cognitive networks that grow over time might be a productive way to further examine typical language development as well as atypical language development (e.g., [41]). We believe the cognitive network approach offers Psychology (and other related disciplines) a simple yet powerful way to model how knowledge is represented in memory, and how the structure of that knowledge might influence cognitive processing.

Author Contributions: Conceptualization, M.S.V. and L.N.-C.; methodology, M.S.V., L.N.-C. and S.N.-C.; software, M.S.V., L.N.-C. and S.N.-C.; validation, M.S.V. and L.N.-C.; formal analysis, M.S.V. and L.N.-C.; investigation, M.S.V. and L.N.-C.; resources, M.S.V., L.N.-C. and S.N.-C.; data curation, M.S.V., L.N.-C. and S.N.-C.; writing—original draft preparation, M.S.V., L.N.-C. and S.N.-C.; writing—review and editing, M.S.V., L.N.-C. and S.N.-C.; visualization, M.S.V., L.N.-C. and S.N.-C.; supervision, M.S.V.; project administration, M.S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data from Simulations 1–3 are available on the Open Science Framework (https://osf.io/MH9JG).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vitevitch, M.S.; Luce, P.A. When words compete: Levels of processing in spoken word perception. *Psychol. Sci.* **1998**, *9*, 325–329. [CrossRef]
2. Vitevitch, M.S.; Luce, P.A. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.* **1999**, *40*, 374–408. [CrossRef]
3. Crystal, D. *A First Dictionary of Linguistics and Phonetics*; Andre Deutsch: London, UK, 1980.
4. Jusczyk, P.W.; Luce, P.A.; Charles-Luce, J. Infants' sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* **1994**, *33*, 630–645. [CrossRef]
5. Saffran, J.R.; Newport, E.L.; Aslin, R.N. Word segmentation: The role of distributional cues. *J. Mem. Lang.* **1996**, *35*, 606–621. [CrossRef]
6. Storkel, H.L. Learning new words: Phonotactic probability in language development. *J. Speech Lang. Hear. Res.* **2001**, *44*, 1321–1337. [CrossRef]
7. Vitevitch, M.S.; Aljasser, F.M. Phonotactics in spoken word recognition. In *The Handbook of Speech Perception*; Pardo, J.S., Nygaard, L.C., Remez, R.E., Pisoni, D.B., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2021.
8. Preston, K.A. The speed of word perception and its relation to reading ability. *J. Gen. Psychol.* **1935**, *13*, 199–203. [CrossRef]
9. Luce, P.A.; Pisoni, D.B. Recognizing spoken words: The neighborhood activation model. *Ear Hear.* **1998**, *19*, 1–36. [CrossRef]

10. Vitevitch, M.S.; Luce, P.A. Phonological Neighborhood Effects in Spoken Word Perception and Production. *Annu. Rev. Linguist.* **2016**, *2*, 75–94. [CrossRef]
11. Vitevitch, M.S.; Luce, P.A.; Pisoni, D.B.; Auer, E.T. Phonotactics, neighborhood activation and lexical access for spoken words. *Brain. Lang.* **1999**, *68*, 306–311. [CrossRef]
12. Vitevitch, M.S. The influence of sublexical and lexical representations on the processing of spoken words in English. *Clin. Linguist. Phonet.* **2003**, *17*, 487–499. [CrossRef]
13. Grossberg, S.; Stone, G.O. Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychol. Rev.* **1986**, *93*, 46–74. [CrossRef] [PubMed]
14. Pitt, M.A.; Myung, J.I.; Altteri, N. Modeling the word recognition data of Vitevitch and Luce (1998): Is it ARTful? *Psychon. B Rev.* **2007**, *14*, 442–448. [CrossRef]
15. Siew, C.S.Q.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive Network Science: A review of research on cognition through the lens of representations, processes, and dynamics. *Complexity* **2019**. [CrossRef]
16. Vitevitch, M.S. *Network Science in Cognitive Psychology*; Routledge: London, UK, 2019.
17. Vitevitch, M.S. What can graph theory tell us about word learning and lexical retrieval? *J. Speech. Lang. Hear. Res.* **2008**, *51*, 408–422. [CrossRef]
18. Steyvers, M.; Tenenbaum, J.B. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cogn. Sci.* **2005**, *29*, 41–78. [CrossRef] [PubMed]
19. McClelland, J.L.; Elman, J.L. The TRACE model of speech perception. *Cogn. Psychol.* **1986**, *18*, 1–86. [CrossRef]
20. Siew, C.S.Q. Community structure in the phonological network. *Front. Psychol.* **2013**, *4*, 553. [CrossRef]
21. Siew, C.S.Q. spreadr: An R package to simulate spreading activation in a network. *Behav. Res. Methods* **2019**, *51*, 910–929. [CrossRef]
22. Valente, T.W.; Dyal, S.R.; Chu, K.H.; Wipfli, H.; Fujimoto, K. Diffusion of innovations theory applied to global tobacco control treaty ratification. *Soc. Sci. Med.* **2015**, *145*, 89–97. [CrossRef]
23. Collins, A.M.; Loftus, E.F. A spreading-activation theory of semantic processing. *Psychol. Rev.* **1975**, *82*, 407–428. [CrossRef]
24. Vitevitch, M.S.; Ercal, G.; Adagarla, B. Simulating retrieval from a highly clustered network: Implications for spoken word recognition. *Front. Lang. Sci.* **2011**, *2*, 369. [CrossRef] [PubMed]
25. Vitevitch, M.S.; Mullin, G.J. *What Do Cognitive Networks Do? Simulations of Spoken Word Recognition Using the Cognitive Network Science Approach*; University of Kansas: Lawrence, KS, USA, 2021.
26. Vitevitch, M.S.; Goldstein, R.; Johnson, E. Path-length and the misperception of speech: Insights from Network Science and Psycholinguistics. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*; Mehler, A., Blanchard, P., Job, B., Banish, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2016.
27. Vitevitch, M.S.; Luce, P.A. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Meth. Ins. C* **2004**, *36*, 481–487. [CrossRef]
28. Aljasser, F.; Vitevitch, M.S. A web-based interface to calculate phonotactic probability for words and nonwords in Modern Standard Arabic. *Behav. Res. Methods* **2018**, *50*, 313–322. [CrossRef] [PubMed]
29. Lumley, T.; Diehr, P.; Emerson, S.; Chen, L. The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* **2002**, *23*, 151–169. [CrossRef]
30. JASP Team. JASP (Version 0.14.1) [Computer Software]. 2020. Available online: https://jasp-stats.org/ (accessed on 16 December 2020).
31. Gow, D.W.; Schoenhaut, A.; Avcu, E.; Ahlfors, S.P. Behavioral and Neurodynamic effects of word learning on phonotactic repair. *Front. Psychol.* **2021**, *12*, 494. [CrossRef]
32. Kello, C.T.; Beltz, B.C. Scale-Free Networks in Phonological and Orthographic Wordform Lexicons. In *Approaches to Phonological Complexity*; Chitoran, I., Coupé, C., Marsico, E., Pellegrino, F., Eds.; Mouton de Gruyter: Berlin, Germany, 2009.
33. R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 5 March 2020).
34. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **2006**, *1695*, 1–9.
35. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
36. Castro, N.; Stella, M.; Siew, C.S.Q. Quantifying the interplay of semantics and phonology during failures of word retrieval by people with aphasia using a multiplex lexical network. *Cogn. Sci.* **2020**, *44*, e12881. [CrossRef]
37. Interdonato, R.; Atzmueller, M.; Gaito, S.; Kanawati, R.; Largeron, C.; Sala, A. Feature-rich networks: Going beyond complex network topologies. *Appl. Netw. Sci.* **2019**, *4*, 1–13. [CrossRef]
38. Citraro, S.; Rossetti, G. Identifying and exploiting homogeneous communities in labeled networks. *Appl. Netw. Sci.* **2020**, *5*, 55. [CrossRef]
39. Dubossarsky, H.; De Deyne, S.; Hills, T.T. Quantifying the structure of free association networks across the lifespan. *Dev. Psychol.* **2017**, *53*, 1560. [CrossRef] [PubMed]

40. Storkel, H.L.; Lee, S.Y. The independent effects of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Lang. Cogn. Proc.* **2011**, *26*, 191–211. [CrossRef] [PubMed]
41. Storkel, H.L.; Hoover, J.R. Word learning by children with phonological delays: Differentiating effects of phonotactic probability and neighborhood density. *J. Commun. Disord.* **2010**, *43*, 105–119. [CrossRef] [PubMed]

big data and cognitive computing

MDPI

Article

A Semantic Web Framework for Automated Smart Assistants: A Case Study for Public Health

Yusuf Sermet [1,*] and Ibrahim Demir [2]

[1] IIHR—Hydroscience & Engineering, University of Iowa, Iowa City, IA 52246, USA
[2] Civil and Environmental Engineering, University of Iowa, Iowa City, IA 52246, USA; ibrahim-demir@uiowa.edu
* Correspondence: msermet@uiowa.edu

Abstract: The COVID-19 pandemic elucidated that knowledge systems will be instrumental in cases where accurate information needs to be communicated to a substantial group of people with different backgrounds and technological resources. However, several challenges and obstacles hold back the wide adoption of virtual assistants by public health departments and organizations. This paper presents the Instant Expert, an open-source semantic web framework to build and integrate voice-enabled smart assistants (i.e., chatbots) for any web platform regardless of the underlying domain and technology. The component allows non-technical domain experts to effortlessly incorporate an operational assistant with voice recognition capability into their websites. Instant Expert is capable of automatically parsing, processing, and modeling Frequently Asked Questions pages as an information resource as well as communicating with an external knowledge engine for ontology-powered inference and dynamic data use. The presented framework uses advanced web technologies to ensure reusability and reliability, and an inference engine for natural-language understanding powered by deep learning and heuristic algorithms. A use case for creating an informatory assistant for COVID-19 based on the Centers for Disease Control and Prevention (CDC) data is presented to demonstrate the framework's usage and benefits.

Keywords: smart assistants; knowledge generation; intelligent systems; web components; deep learning; web-based interaction

Citation: Sermet, Y.; Demir, I. A Semantic Web Framework for Automated Smart Assistants: A Case Study for Public Health. *Big Data Cogn. Comput.* **2021**, *5*, 57. https://doi.org/10.3390/bdcc5040057

Academic Editors: Massimo Stella and Yoed N. Kenett

Received: 8 June 2021
Accepted: 12 October 2021
Published: 18 October 2021

1. Introduction

Rapid advancements in monitoring and computational techniques have led to the abundance of semantically connected and annotated data in a plethora of fields [1,2], thus presenting the need for effective tools for its management [3], analysis, and communication [4]. Web-based information systems (IS) serve as one-stop platforms to access, analyze, and explore information effectively for decision-making purposes [5,6]. Although current systems are proved to be successful to communicate and analyze data efficiently, they still suffer from the complexity and the higher learning curves due to the limitations of conventional interaction methods [7]. Users of information systems (e.g., public, workers, managers, decision-makers, organizational leaders) often look for a certain piece of knowledge for which they may have to master the functionalities and resources provided by the IS [8]. This is especially tedious and discouraging for users who are not continuous visitors to the system. Thus, modern approaches are needed to free the users from the nuances and complexities of information systems and provide a feasible tool to access knowledge [9]. Automated knowledge communication will be critical for participatory decision-making process in disaster mitigation as well [10,11].

With the advancements in artificial intelligence, there is a significant surge in research of chatbots, which can be defined as intelligent agents (i.e., assistants) that can comprehend natural-language queries and produce a direct and factual response using data and service providers [12]. More recently, the increased prevalence of deep learning tools and

generalized algorithms resulted in the growth of chatbots fueled from deep learning's ability to make robust and scalable inferences out of high-dimensional and likely noisy data [13]. Technology companies have been taking the lead on operational virtual assistants integrated into their ecosystem which triggered a brand new and massive market that was valued at US$ 2.2 billion in 2018 and is expected to reach US$ 11.3 billion by 2024 [14]. However, the usage of chatbots for effective and reliable information communication is not widespread among public, government, and scientific communities [15]. As of 2018, only 16% of Internet users have ever used a chatbot [16] even though there is a substantial demand [17]. Several publications emphasize the potential chatbots hold to serve as the next generation information communication tool and make the case for an urgent need for chatbot development and adoption in their respective fields [18]. The United States Army Corps of Engineers (USACE) Institute for Water Resources reports that virtual assistants hold great promise for the USACE to better serve the public, enhance workflow, and improve decision-making while saving time and financial resources [19]. Voice-enabled knowledge systems can support novel virtual reality systems [20] for emergency response training and disaster risk communication [21]. More recently, the rise of the COVID-19 pandemic made it clear that chatbots will be instrumental in cases where accurate information needs to be communicated to a substantial group of people with different backgrounds and technological resources as reported in a Nature article [22]. Sohrabi et al. [23] describe how artificial intelligence-powered chatbots can be widely used to provide up-to-date information regarding the COVID-19 outbreak in the context of prevention and management.

Relevant Work

Conversational intelligent agents have been extensively researched and applied in many fields, including health care [24], finance [25], environmental sciences [26], psychology [27], and administration [28]. They have required the development of taxonomy, natural-language processing approaches, interaction methods and interfaces, and intent mapping (e.g., scoring), and are often developed for specific use cases by an experienced developer team. More recently, commercial chatbot development (e.g., Microsoft Bot Framework, Google DialogFlow, IBM Watson, Facebook Wit.ai, Amazon Lex) services have been made available to streamline the chatbot development workflow by separating the functionality from the use cases and providing on-demand modular services for the building blocks of a smart assistant (e.g., intent classification, voice recognition, speech synthesis, cloud computing for query execution) [29]. However, these services are often costly and black-box, and may be limited to their own ecosystem, which may result in being bound to cloud services of the respective software ecosystem [30]. Furthermore, such paid platforms often come with a variety of advanced features that may not be needed by all use cases (e.g., integration to messaging applications). An exception is RASA, a commercial and partially open-source conversational AI platform to develop chatbots. Although its open-source natural-language understanding module and integration connectors allow developers to build and host chatbots, voice recognition, synthesis, and web tools for encapsulated UI/UX are not provided within the framework [31]. Similar to RASA, another recent chatbot development framework that is commercial and open-source is Xatkit. Although its features and limitations are partially similar to RASA, its major distinctions include e-commerce support and increasing the abstraction of chatbot development process [30]. Developing full-scale conversational agents are difficult to develop and maintain even by using commercial products, due to the sheer complexity of achieving human-like interaction (e.g., detecting meaning, maintaining a context, handling unforeseen situations meaningfully) [32]. The budget and the talent it requires to have such agents make it unfeasible for many educational institutes, research organizations, and micro-enterprises to benefit from natural-language interactions.

The futuristic vision of widespread employment of virtual assistants is held back by several challenges [33]. One of the major obstacles of virtual assistant development and

use is that the available services usually require a centralized data flow (e.g., server-side processing of voice, sentence, or data) involving a third-party service provider which presents concerns in terms of data privacy, security, and compatibility [30]. Currently available frameworks for chatbot development often employ vertical integration in which the developers find themselves stuck with the ecosystem of the service provider [34]. A 2019 report on the virtual personal assistant market concludes that the high-growth inflection point of the market and wide adoption of smart assistants in many fields will occur as open-source and free assistant development tools [34]. A second major challenge is that the workforce and/or financial resources required to initialize and maintain a chatbot discourages establishments, governments, research groups, and non-profits, while for larger companies and organizations these costs justify themselves for the long term [35]. Furthermore, active internet connection is often needed for chatbot to operate at its simplest level, which presents a significant setback as many use cases do not have a reliable and consistent internet connection [19].

Thus, while the literature review reveals the need and a plethora of areas of research for virtual assistants, the research questions (RQ) that this study aims to tackle are formulated as follows.

1. Can natural-language question answering be democratized for under-resourced organizations and teams?
2. Can a voice-enabled smart assistant be realized and function with client-side computational resources?
3. Is it possible to ensure data privacy while using voice assistants?
4. Can a caching mechanism be employed for offline chatbot usage?

This paper presents Instant Expert, an open-source semantic web component to build and integrate voice-enabled smart assistants (i.e., chatbots) for any web platform regardless of the underlying domain and technology. The component allows non-technical domain experts to incorporate an operational assistant with voice recognition capability into their websites by simply adding as little as a single line of HTML code while customizations are enabled for more advanced use cases. The component entails an encapsulated user interface that accepts natural-language questions via text and speech inputs as well as selection from a predefined list of questions. A knowledge generation module processes questions to map them to the configured data resource and returns the answers using its inference engine and natural-language mapping methods powered by deep learning and heuristic algorithms. Instant Expert is capable of automatically parsing, processing, and modeling internal (same-origin) or external (cross-origin) Frequently Asked Questions (FAQ) webpages as an information resource as well as communicating with an external knowledge engine for more advanced use cases. The presented framework is powered by advanced web technologies to ensure reusability and reliability and deep learning to perform accurate natural-language mappings. A use case for creating an informatory assistant for COVID-19 based on the Centers for Disease Control and Prevention (CDC) data is presented to demonstrate the framework's usage and benefits.

The main contributions of this research can be summarized as follows. The presented component makes it possible for any web system on any domain to have its own voice-enabled smart assistant to instantly provide factual responses to natural-language queries. It can grow the system's visibility and increase user retention and satisfaction due to providing the user with the information they desire without a hassle. The component can especially be valuable for individual developers, academic research groups, small companies, non-profits, and public offices that may not have the resources for the development and maintenance of commercial smart assistants for their organization. The framework liberates developers from the limitations and boundaries of any given ecosystem and maximizes customization based on available resources and needs while providing a generalized framework to offer standardized, robust, and efficient smart assistants. The framework is completely built on web technologies (e.g., HTML5, JavaScript, CSS) working on the client-side which eliminates the dependence to server-side technologies and assures

data privacy. Finally, the Instant Expert's modular architecture, which is not bounded by any ecosystem, paves the way for its expansion into virtual reality (e.g., A-Frame), and augmented reality (i.e., HoloLens and Magic Leap) applications through web [36].

The remainder of this article is organized as follows. Section 2 presents the methodology of the development and implementation of the intelligent web framework. Section 3 describes a case study on public health and provides benchmark results and performance analysis. Section 4 concludes the articles with a summary of contributions and future work.

2. Methods

The Instant Expert web component entails several components all of which are implemented using Hypertext Markup Language 5 (HTML5), JavaScript (JS), and Cascading Style Sheets (CSS). The component runs completely on the client-side (i.e., using only the client's hardware) by which significantly minimizes the workload, and consequently, the maintenance cost to employ a smart assistant. The framework can be abstracted into two main semantic units: (1) user interaction and interface and (2) knowledge generation. Various features of ECMAScript 6 (ES6) have been used in the software including JS Modules to avoid global namespace pollution and provide a degree of encapsulation against potential conflicts between the smart assistant and hosting web platform. The framework can be imported into any web application as a script as demonstrated below.

```
<script src="instant-expert-dist.js"></script>
```

Additional benefits of using ES6 include default parameter values, string interpolation, OOP-style (Object-Oriented Programming) classes, which lead to a robust, clean, and extendable product. The usage of web frameworks (e.g., Polymer, Stencil) is considered when designing the web component, though consciously avoided to minimize the learning curve and eliminate the dependency on any framework's abstraction. The code of Instant Expert is fully open-source and accessible via GitHub (https://github.com/uihilab/instant-expert, accessed on 11 October 2021) with detailed documentation guiding developers on how to employ and configure the framework. Figure 1 summarizes the overall architecture of the presented framework.

2.1. User Interaction and Interface

The Instant Expert is built upon the features provided by the Web Components. Web Components are a collection of web technologies combined with the purpose of creating reusable, customizable, and encapsulated HTML elements [37]. It is mainly powered by three web technologies. Custom Elements enables the creation of new HTML elements. HTML templates offer the mechanism to define HTML content that can be instantiated during runtime instead of being rendered when the page is loaded. Shadow DOM provides encapsulation of an element's features with a shadow tree associated with the web component [38]. The major consideration when designing the component is to prevent it to affect the visual and functional integrity of the hosting web site, as well as to prevent it from being affected by the hosting web site. Shadow DOM prevents potential integration complications by creating a shadow root under the custom HTML element and rendering it separately from the main document DOM. These new technologies are not yet widely adopted in the industry and especially in academia despite the remarkable benefits they offer. Another aim of this paper is to serve as an example to advantages of Web Components by reducing the technicality.

CSS Containment

Containment is a CSS feature that aims to isolate a contained element's contents from the rest of the document, as much as possible. The main purpose of the containment is to provide optimization and offer stability in rendering and painting of web pages by providing a standardized way. Browsers are always looking to make optimizations when rendering a page. Containment provides a standardized way to tell the browser where and how it can optimize without breaking the intended functionality. When using third-party

DOM, such as the Instant Expert, containment can prove to be useful to sandbox the component to protect and increase the performance of the page. It should be noted that containment is not a security feature and is not aimed to provide a full encapsulation.

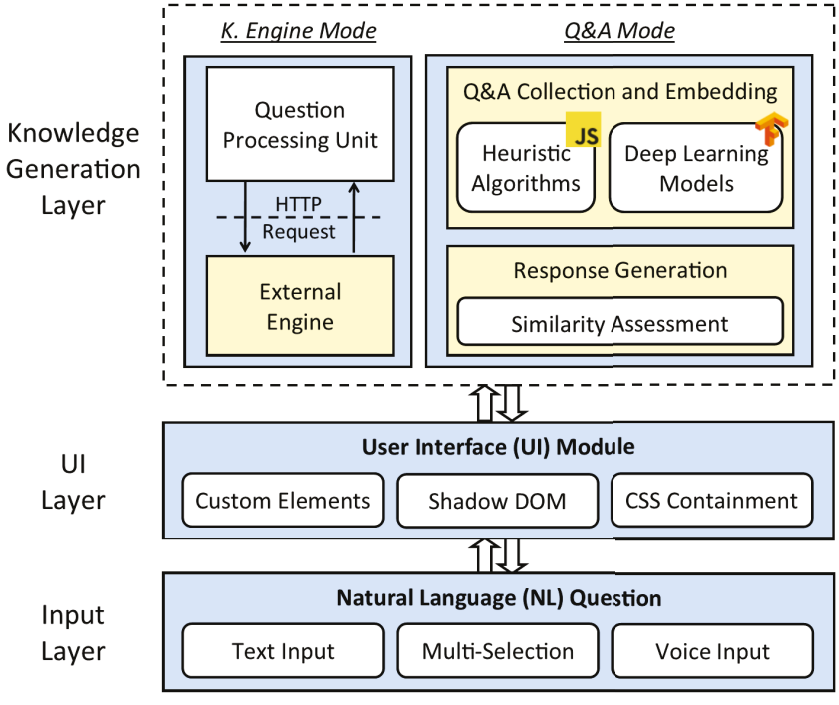


Figure 1. System architecture of the Instant Expert.

There are 4 main types of containment that can either be used individually or in groups. The details of each containment type can be found in the specification document published by the World Wide Web Consortium (W3C) [39]. For the purposes of Instant Expert, Content Containment has been used which combines Layout, Paint, and Style Containments. A web component already brings the containing functionality; however, Content Containment introduces significant performance benefits and decreases rendering runtime. This is especially valuable to protect the performance of the website that integrates the Instant Expert. As of June 2020, CSS Containment is supported by default by the latest versions of major browsers (i.e., Google Chrome, Opera, Mozilla Firefox Nightly, and Microsoft Edge). Another advantage of using Content Containment is that it will not affect the functionality of Instant Expert even if the client browser does not support it.

User Input

There are three ways for a user to interact with the system; (a) manually typing the question to a text box, (b) invoking voice recognition to ask the question using a microphone, and (c) selecting from a predefined list of questions (Figure 2). Text input is the most common type of interaction due to the search engine culture. Having a predefined list of questions allows the user to explore the system and better understand its capabilities. Voice-enabled communication is supported using Web Speech API, which is an experimental technology that defines a JavaScript API to integrate speech recognition and speech synthesis functionality into web pages. As of June 2020, speech recognition is supported by the latest versions of Google Chrome, Opera, and Microsoft Edge browsers. Speech synthesis is supported in all major browsers (e.g., Google Chrome, Mozilla Firefox, Opera, Microsoft Edge, and Safari). The component checks the client browser at initialization to

test if Web Speech API is supported and disables speech features if not supported. The component allows the incorporation of third-party speech recognition and synthesis APIs; however, it requires modification of the component's source code. Upon the construction of the natural-language question in text format, all input types eventually follow through the same flow to be passed to the knowledge generation module.

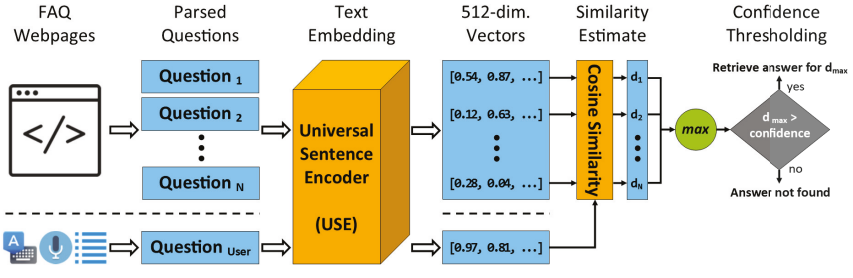


Figure 2. Flowchart summarizing the Q&A Mode's execution logic.

2.2. Knowledge Generation Module

The knowledge generation module is tasked with the complete workflow of producing a direct response to the given natural-language question. There are two main approaches for powering the Instant Expert with a knowledge base; knowledge-engine mode and question-and-answer (Q&A) mode. The knowledge-engine mode relies on the integration of an external server-side knowledge engine that can access and analyzing vast amounts of data in response to a natural-language question. The Q&A mode requires a list of question-and-answer pairs supplied manually or by providing the URL of a webpage containing such pairs (e.g., Frequently Asked Questions). The Q&A mode uses deep learning models to create a tensor representation for each question in the knowledge base to calculate semantic similarity with the user question.

2.2.1. Q&A Mode

A major motivation of Instant Expert is to augment existing web platforms with a plug-and-play importable web component with minimal effort. In most use cases, static and textual responses can suffice to help users find useful information that they were looking for. Such pieces of information are often presented in a web platform in the form of Frequently Asked Questions (FAQ). However, searching for information via FAQs is often discouraging, hard to navigate, time-consuming, and results in failure of communicating critical information in a timely manner [40]. As a solution, the Q&A Mode of the Knowledge Generation Module is equipped with functionalities to parse and process FAQ pages and to efficiently and effectively map any user question into one of the question-and-answer pairs in the generated knowledge base. Thus, the Instant Expert effortlessly enhances user experience by allowing the users to verbally communicate with the system and receive a direct response without the hassle of going through potentially hundreds of frequently asked questions. Figure 2 visualizes the workflow for the Q&A Mode.

Q&A Parsing and Collection

There are two approaches to collect Q&A pairs: manual entry and automatic FAQ parsing. For web platforms that do not have already established FAQ pages, the web component allows the developers to enter questions and their respective answers through HTML <slot> elements within the scope of Instant Expert in a hierarchical and organized manner. Since the Q&A pairs are directly provided, no further processing has been done before the encoding stage.

Automatic FAQ parsing requires the developer to supply the direct URL of the web-page containing the FAQ into the web component as an attribute. Both same-origin and cross-origin addresses are accepted given that the FAQ pages, which are not hosted in

the same domain as the application, are navigated via a proxy server. The strength of the FAQ parsing and Q&A extraction comes from its unique design where no preprocessing is required regarding the webpage structure. Heuristic algorithms can perform just as well or better than machine learning approaches for similar scraping tasks as reported in the literature [41]. Our novel heuristic algorithm is built based on two assumptions concerning the structure of the FAQ webpage: (1) most question mark characters belong to the FAQ on a given page, and (2) both the questions and answers follow a certain HTML structure (i.e., not arbitrarily placed). The first assumption is clearly the case for the majority, if not all, of the FAQ pages since most FAQ questions end with a question mark and it is extremely unlikely to have more question marks in the remaining of the page. As for the second assumption, all questions in a given FAQ webpage is highly likely to follow a certain convention for the purpose of clarity, symmetry, maintainability, and effectiveness. More concretely, the HTML elements containing the FAQ question will likely share certain characteristics such as element tag, class, styling, immediate parents, and DOM depth. Thus, a pattern can be inferred by analyzing these correlations in an efficient way. The same condition is applicable for the FAQ answers.

The parsing process begins with finding innermost HTML elements containing a question mark to retrieve unique texts. Per our assumption, most of these question marks should belong to FAQ questions. To deduce a pattern, we traverse each retrieved element to keep record of their and their closest three parents, if exist, tag names along with their depth within the DOM tree. Per our symmetry assumption, the elements containing FAQ questions should produce the same values for tag names and depth. Using a hash table, the frequency of each combination is measured to identify the correct pattern. Thus, all FAQ questions are successfully retrieved based on the inferred pattern. Before initiating the process for parsing the FAQ answers, a challenge arises regarding the scope. For example, the innermost element containing the question might have parent elements for visual and interactive purposes (e.g., buttons, container boxes) before continuing to the following question or answer. It is known that the answer to a question lies somewhere between the text of that question and the text of next question. However, it cannot be known whether the answer is grouped with its question:

```
<div id="questionGroup1">
    <div id="question1Container">...</div>
    <div id="answer1Container">...</div>
</div>
```

or simply have been coded sequentially in the same scope with other questions and answers:

```
<div id="question1Container">...</div>
<div id="answer1Container">...</div>
```

Even if processing would have been performed to find the uppermost parent containing only the question text, scope issues still would not fully resolve since it is a common practice in FAQ pages to group questions by topic, which breaks the sibling relationship between the last question in a group and the first question in the next group. As a solution, a heuristic approach is taken to deduce a pattern to work in all possible webpage configurations. First, the uppermost parent of each question that does not contain its following question is retrieved, and its distance to the child is recorded. Per the symmetry assumption, most retrievals will share the same distance value. Thus, the system can extract all answers regardless of their scope and structuring. Thus, all Q&A pairs are extracted from an FAQ webpage in a heuristic manner requiring no data other than the URL of the Q&A page.

Q&A Encoding

The extracted Q&A pairs are processed to create an encoded and numeric representation of the questions so that the user question can effectively be mapped. Due to the high

variety of expression styles for questions sharing the same intent and desiring the same output, the structure and the words in the sentence should be analyzed while preserving the semantic integrity [42]. Furthermore, this representation will only be used to calculate similarity between different question sentences, and thus, natural-language parsing and data extraction is not a necessity. An additional design choice is to be able to efficiently perform the encoding process on the client-side. Given these requirements, the Instant Expert uses the Universal Sentence Encoder (USE) [43], a model that is designed to embed text to produce a vector representation entailing the semantic information within the sentence. More specifically, the Transformer architecture [44] constitutes the underlying encoding model for context-aware representations of words resulting in a 512-dimensional vector. Within the presented framework, a lightweight version of the USE with an 8000-word vocabulary is imported and executed with TensorFlow.js, an open-source JavaScript library that can train and deploy machine learning models on the client-side via the browser.

All questions (n questions) in the collected Q&A pairs are provided as input to the USE model via TensorFlow.js which produces a 512-dimensional tensor per question and results in a [n, 512]-sized matrix. This process takes place asynchronously in the background while allowing the users of the website to continue their interaction as usual. Once the processing is complete, the tensor matrix as well as the list of Q&A pairs are saved within the duration of the session. If the developer enabled *downloadModel* attribute of the Instant Expert web component, then the framework will generate a JSON file consisting of the tensor matrix and the Q&A pairs. This JSON file can be hosted on a server and the URL to access the file can be provided to Instant Expert. This mechanism allows the developers to prevent the client's hardware resources to unnecessarily be used by the embedding process, and make the Instant Expert instance to be initialized instantly to respond to any questions without needing to wait for the asynchronous operations to be completed. Thus, the framework allows three different usage styles for the Q&A Mode: same-origin or cross-origin FAQ webpage, manual Q&A definition, and JSON file containing Q&A pairs and question embeddings.

Response Generation

Each time a user asks a question through Instant Expert, the USE model is used in the same fashion to generate the embeddings in the form of a tensor. Thus, every predefined question with known answers as well as the input question are represented as points on a 512-dimensional coordinate system. Distance between these points is used as the criterion to assess semantic similarity. As the distance measurement technique, this framework uses the Cosine Similarity (Equation (1)) which is defined as;

$$\cos\theta = \frac{X \cdot Y}{||X||\ ||Y||} \tag{1}$$

where X and Y are the tensor representations of the user's question and a predefined question in the parsed Q&A pairs, respectively. For robust calculation of dot products and norms, an open-source JavaScript library (Math.js) is used. This similarity is calculated for n questions that result in a similarity array where the values closer to 1 indicate higher correlation. Before completing the mapping to the question with the highest similarity, a threshold value is established to filter unrelated questions where a satisfactory answer is not present in the knowledge base. This is especially important to eliminate the potential to misguide users with unrelated or unappropriated information. This threshold value can be adjusted according to the nature of the use case to determine the trade-off between precision and recall. The framework performs the response generation phase asynchronously so that the hosting website is not throttled or disturbed. The generated response is returned to be displayed via the web component's interface along with the source of information to ensure the recipient of the information is aware of the source.

2.2.2. Knowledge-Engine Mode

The presented component can be connected to an external knowledge engine for more sophisticated use cases of intelligent assistants. A server-side engine may use semantic webs to aid in the inference process and dynamic and distributed data resources to respond to complex queries. The main advantage of such engine is the use of custom natural-language processing approaches capable of extracting useful information from the question such as time and date, location, intent, output type (e.g., graph, image, numeric value), ontological entities, and mathematical and statistical operations [45]. Thus, the knowledge generation process pertaining to an external engine has a computational essence in comparison to merely providing textual information. Furthermore, this mode provides the option for users to use existing open-source conversational AI development frameworks (e.g., RASA, Xatkit) for insightful interactions.

The web component retrieves the answer for the input question by making an HTTP POST request to an external knowledge engine using the webhook link provided with the 'engine' attribute of the element. The only parameter passed to the engine is the question text using the parameter key 'question', which can be modified per developer's configuration. The component expects a response from the engine in JSON format with a key-value pair where value represents the response. The module requires the response to return within 2 s to portray a realistic and natural interaction with the user. If the engine and the web page that integrates the Instant Expert are not hosted on the same origin, then the engine should be able to handle requests from origins outside of its own by setting up Cross-Origin Resource Sharing (CORS), or use other solutions such as to use a proxy server or script.

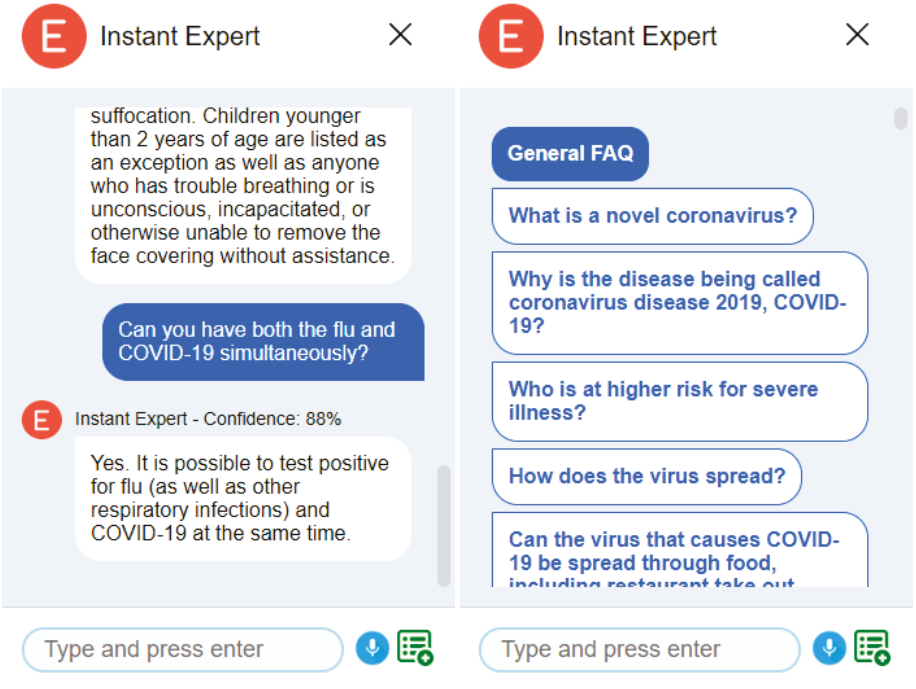
3. Results

3.1. Q&A Mode–Public Health Case Study

The generalized nature of the presented framework makes it suitable for usage in any domain. For the purpose of demonstrating its workflow and benefits, an example use case needs to be selected where the necessity and impact of chatbots are evident. The degree of misinformation and sparseness of sources results in a data mess for topics ranging from politics to health. Social media and online news outlets often receive the same information from the same source, but report it as duplication through a story with a narrative, which in turn, results in information pollution. One of the advantages of the Instant Expert is that it allows the sharing of information through numerous challenges by referencing the original source instead of duplication which opens the way for distrust. A recent example of such scenario is the ongoing pandemic of Coronavirus Disease 2019 (COVID-19). It has been widely reported how chatbots can prove to be useful during the pandemic on different levels envisioned and developed by various organizations and companies. Examples to these chatbot use cases regarding COVID-19 include, but not limited to, assessing eligibility for plasma donation [46], symptom checkers and health screening [47–49], and information dissemination, which is especially crucial as chatbots provide a direct and single response to a given question instead of the alternative of being succumbed to social media posts and the spread of misleading or incomplete information [22,50]. It has been advised that local authorities and healthcare businesses should use chatbots to ensure 7/24 accurate information flow powered by the extensive amounts of Frequently Asked Questions available by authoritative sources such as CDC and WHO [51]. Yet, the realization of this vision is hindered by the lack of technological capabilities, resources, and funding available to such local and healthcare organizations. Thus, an information dissemination chatbot for COVID-19 has been selected as the use case in this paper to demonstrate the presented framework's usage and benefits due to the urgent demand as COVID-19 pandemic is progressing.

According to web analytics service [52], the CDC website has received the highest number of visits (i.e., traffic) among websites that are served in English and that offer information and statistics on the spread of the COVID-19 infection. Thus, we have selected

the CDC's official Frequently Asked Questions webpage [53] as a source for the following use cases. On that page, there is a total of 119 questions spanning various topics ranging from COVID-19 basics to cleaning and disinfection as of 20 June 2020. Figure 3 shows screenshots from the Instant Expert instance with CDC-powered Q&A set.



(**a**) (**b**)

Figure 3. Screenshots from the Instant Expert for COVID-19 Chatbot. (**a**) The main screen of conversation with buttons to activate voice input and display example questions, (**b**) example questions that can be asked to the chatbot as recommendations to the user.

The case study has been performed on an average personal computer that is powered by Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz, 32 GB 2400 MHz RAM, and an NVIDIA GeForce GTX 1050.

3.1.1. FAQ from a Web Page

The Instant Expert has been initialized in the FAQ mode and provided the CDC's FAQ URL as the information source. Due to the cross-origin limitations, a proxy server (CORS Anywhere) has been used to retrieve the webpage contents. Exactly all 119 Q&A items on the CDC page have successfully been extracted with a 100% precision and recall, followed by the embedding of all questions for natural-language mapping. The average time spent on parsing the Q&A pairs and for embedding the questions using the Universal Sentence Encoder has been reported in Figure 4. To visualize the processing complexity, this benchmark has been performed for subsets of the CDC questions so that the correlation between the number of questions and the processing time is clear. The runtime measurement does not include the time spent for the retrieval of the FAQ page via GET request and loading the USE model as both of their performance is out of the presented framework's scope and affected by external parameters such as the internet speed. Furthermore, these actions

are only performed once, and thus, is neither related to the number of questions nor the structure of the FAQ page.

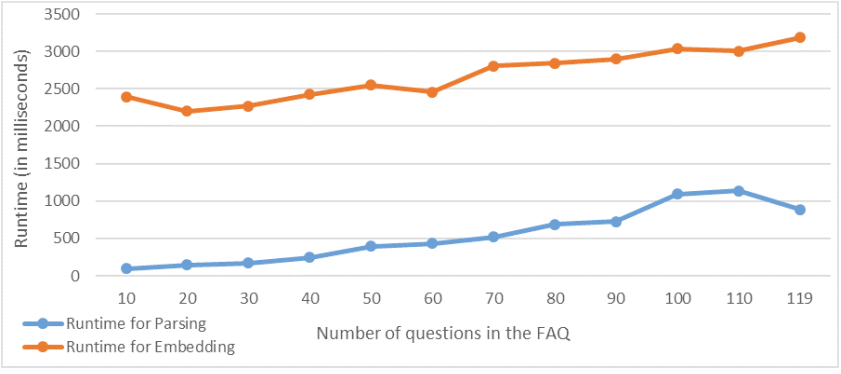


Figure 4. Benchmarks for powering the Instant Expert from an FAQ webpage: parsing time and embedding time.

One of the main purposes of the FAQ mode is to allow users to ask questions in various natural-language expressions. A test set is generated to quantify how flexible the Instant Expert is in terms of accurately mapping question variations that share the same intent. The test set contains the original FAQ question, answer, and one natural-language question that expects the same answer with different phrasing. For objectivity, a third-party software (i.e., QuillBot), which is a machine learning-powered paraphraser and sentence restructurer, is used to produce high-variance natural-language questions with a similar meaning to the original. Minor manual modifications have been made to questions to challenge the inference component of the system by referencing the actual questions asked by people on the Internet. Table 1 provides examples of the generated questions in comparison to the original question. Additionally, the test set also contains three questions that the CDC's FAQ cannot and should not answer. These three questions are taken from the US Federal Drug Administration's (FDA) FAQ webpage. The reason for this addition is to make sure that the chatbot does not disseminate inaccurate and unrelated information. Thus, the test contains a total of 122 questions. For measurements, a benchmark analysis has been carried out to experiment with a broad range of confidence threshold values with respect to the precision and recall values. For reference, precision represents the percentage of questions that have been mapped to the correct answer out of all the questions that are mapped to some answer. Recall, on the other hand, represents the percentage of questions that have been accurately mapped. Both the benchmark code and a complete test set can be found on the GitHub repository for reproducibility and reanalysis.

To quantify the model's effectiveness, precision (Equation (2)), recall (Equation (3)), and f1-score (Equation (4)) metrics have been selected for this imbalanced classification problem with multiple classes as formulated below [54]. n value in the equations below represents the number of different questions in the FAQ (i.e., classes).

$$precision\ (multiclass) = \frac{\sum_0^n TruePositive}{\sum_0^n TruePositive + \sum_0^n FalsePositive} \tag{2}$$

$$recall\ (multiclass) = \frac{\sum_0^n TruePositive}{\sum_0^n TruePositive + \sum_0^n FalseNegative} \tag{3}$$

$$f1 - score\ (multiclass) = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

Table 1. Example Questions from the Test Set.

Original Question	Generated Test Question
"Should I make my own hand sanitizer if I cannot find it in the stores?"	There are no hand sanitizer left in stores. Should I make one myself?
"What should I do if there is an outbreak in my community?"	What are you suggesting me to do if my community suffers from an outbreak?
"Should I go to work if there is an outbreak in my community?"	Am I supposed to continue working if we have an outbreak in my street?
"Can CDC tell me or my employer when it is safe for me to go back to work/school after recovering from or being exposed to COVID-19?"	If I am exposed to COVID-19, when can I safely go back to work?
"My family member died from COVID-19 while overseas. What are the requirements for returning the body to the United States?"	What is the policy on bringing my relative back to US who passed away due to COVID-19?
"What is routine cleaning? How frequently should facilities be cleaned to reduce the potential spread of COVID-19?"	How often should I clean my place to prevent COVID-19?
"What should I do if there are pets at my long-term care facility or assisted living facility?"	What steps I should take if my nursing home has pets?
N/A (CDC's FAQ does not have this question)	Am I at risk of serious complications from COVID-19 if I smoke cigarettes?
N/A (CDC's FAQ does not have this question)	Are there any vaccines to prevent COVID-19?
N/A (CDC's FAQ does not have this question)	Are antibiotics effective in preventing or treating COVID-19?

Figure 5 shows how precision and recall values are affected based on the selected confidence threshold. To achieve an accurate and complete system, the goal is to maximize the precision and recall values; however, a trade-off evaluation is required [55]. Based on the use case, it may be important to maximize the precision for making sure to always provide the accurate answer for questions that are mapped or to maximize recall for mapping as much question as possible while potentially sacrificing the accuracy of a few. For this use case, the recall starts to drastically decrease around confidence level of 0.8 while the precision stabilizes as seen in Figure 5. Looking at a confidence level of 0.75, the precision is maximized with a value of 100% while recall value is 97.541%. Coincidently, the ideal confidence value for the highest possible recall is also observed to be 0.75, though for other cases this value might differ resulting in a trade-off to prioritize either maximum precision or recall. This ideal confidence level can also be observed with the harmonic mean of precision and recall values (i.e., f1-score) where f1-score is the highest at 0.75 confidence.

For applications when the prediction results in critical information regarding human health, such as the use case for COVID-19, precision is more critical than recall, because providing the wrong information can have detrimental results. Thus, based on the experiment, the confidence threshold is set to 0.75 by default to ensure perfect precision of 100%, which results in 3 out of 119 questions are left unmapped. These three questions are examined to identify why they have not reached to satisfactory confidences. Table 2 summarizes the original FAQ text, the test question, and commentary on why the mapping was not successful. Based on the investigation, we conclude the questions that are not context-independent and not well-structured are more difficult to be mapped from to-the-point questions that an average user of the system might ask.

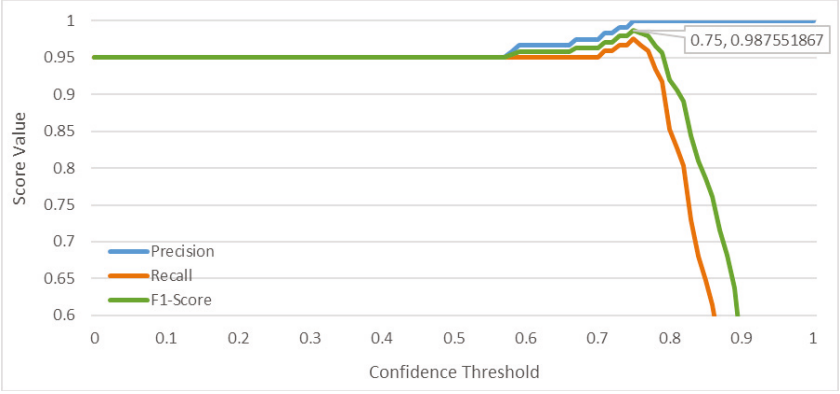


Figure 5. Precision, recall, and f1-score values for different confidence thresholds based on generated test data.

Table 2. Summary of unmapped questions.

Original Question	Test Question	Commentary
"Limit time with older adults, including relatives, and people with chronic medical conditions."	"Should I avoid spending time with the elderly, especially those with health conditions?"	The structure of the original question is not in the form of a question, and rather a recommendation. Especially since there are similar questions exist in the FAQ on dealing with people with underlying conditions, the mapping process could not complete with a satisfactory confidence.
"Will businesses and schools close or stay closed in my community and for how long? Will there be a "stay at home" or "shelter in place" order in my community?"	"For how much longer the business will stay closed?"	This FAQ item is too broad, and in fact entails three different questions for business or schools. The test question only asks a portion of the FAQ item concerning the businesses, which results in unsatisfactory confidence.
"What about imported animals or animal products?"	"Do animal products pose risk?"	This FAQ question is incomplete as its meaning depend on a previous question in the FAQ list.

The Instant Expert can be instantiated based on a FAQ webpage by simply providing the URL. This use case is especially suitable for cases where the information in the FAQ changes frequently. As an example, it has been widely reported that the constantly and rapidly changing nature of COVID-19-related data and guidelines cause misinformation or confusion of the public [56]. Thus, this mode can justify the required client-side workload when the source is an ever-expanding and ever-changing list of question and answers. Example usage is shown below.

```
<instant-expert
    mode="faq-web"
    faq-url="YourProxyURL/ https://www.cdc.gov/coronavirus/2019-ncov/faq.html">
</instant-expert>
```

Parsing FAQs from Other Sources

As a tangent to the COVID-19 Case Study that is powered by CDC data, we have experimented with FAQ webpages from other websites and domains to showcase the Instant Expert's generalized structure and domain independence. Table 3 summarizes five FAQ sources that were used in the same manner as described in the implementation above by providing the respective URL. The output of the parsing component has been recorded and analyzed to measure the success rate. In this context, precision represents the number of the accurate text groups that were labeled as a question or an answer, whereas the recall represents out of all the question and answers on the given webpage the framework was able to accurately parse. The analyzed FAQ webpages are structured in various ways and in different languages (e.g., English, German, Spanish, French), thus, demonstrating the parsing process' generalized workflow. All FAQ pages below analyzed as of 26 June 2020.

Table 3. Analysis of FAQs to demonstrate the framework's domain independence and generalized nature.

FAQ Source	No of Q&A	No of Parsed Q&A	Precision	Recall
FDA COVID-19 FAQ (fda.gov)	78	78	100%	100%
World Health Organization (WHO)—Q&A on coronaviruses (who.int)	24	24		
United Nations COVID-19 FAQ (un.org)				
(in English)	40	40		
(in French)	37	37		
(in Spanish)	38	38		
Stanford COVID-19 FAQ (stanfordhealthcare.org)	16	16		
Robert Koch Institut COVID-19 FAQ (rki.de)				
(in German)	43	43		

3.1.2. FAQ from a Custom List

Some use cases may require manual definition of question and answers instead of having or relying on an existing FAQ webpage. To enable such initialization, the Instant Expert presents a mode, called FAQ-custom, in which HTML Slot elements are used to allow the developer to specify questions and their corresponding answers as shown below.

```
<instant-expert mode="faq-custom">
    <div slot="questions">
        <p>Question 1</p>
        <p>Question 2</p>
    </div>
    <div slot="answers">
        <p>Answer 1</p>
        <p>Answer 2</p>
    </div>
</instant-expert>
```

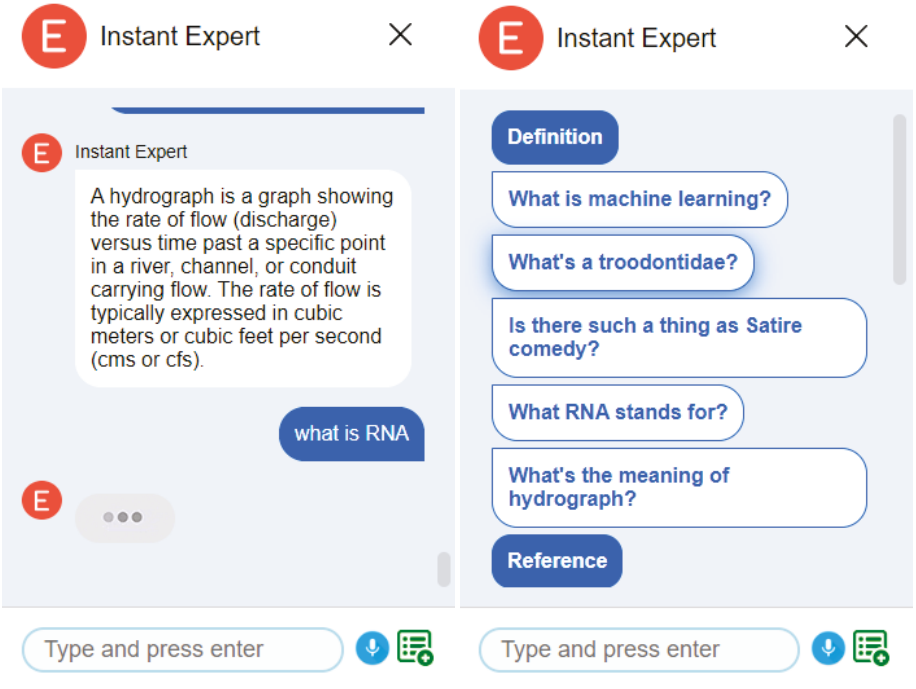
3.1.3. FAQ from a Model

Most of the parsing and embedding process takes place on the background (i.e., async) to allow users to continue normal operation; however, it still consumes client resources and requires varying time depending on client hardware. Both previous FAQ processing approaches (i.e., web, custom) come with the capability of extracting a JSON file containing the parsed Q&A pairs along with their USE embeddings (e.g., a 512-dimensional tensor). This downloaded model file can be provided to the Instant Expert directly to eliminate the time and resources required for FAQ processing. This use case is suggested as the default method to ensure the users can use the chatbot immediately after the page is loaded. Since no new processing is done, the precision and recall values are the same as reported above. Example usage is presented below.

```
<instant-expert
    mode="faq-model"
    faq-url="instantexpert_faq_cdc.json">
</instant-expert>
```

3.2. Knowledge-Engine Mode-Web Search Use Case

An example use case of the presented framework in Knowledge-Engine Mode is the development of a generic question-answering assistant using the Project Answer Search by Microsoft Cognitive Labs. Project Answer Search is an experimental technology to instantly answer natural-language user queries with factual responses [57]. This use case is specifically developed to be a complete solution that (a) will serve as an adoption guide to the users, (b) can be easily accessed to try the presented software, (c) and can be conveniently reproduced for production use or ensure the component's correctness. There are two parts constituting this use case; developing a question-answering engine as the backend and developing a website that implements the Instant Expert component integrated with the engine. The backend is implemented as a Node.js application and served on a cloud platform (i.e., Heroku). The source codes for both the backend and the website are available on Instant Expert GitHub repository in the example directory along with the directions necessary for reproduction. The validation of use cases of the Knowledge-Engine Mode belongs to the external natural-language question-answering service where the accuracy can be measured for response generation. For this use case, an already validated system (i.e., Project Answer Search) is used to focus on the framework's communication mechanism. Figure 6 shows this use case in action.



(**a**) (**b**)

Figure 6. Screenshots from the Instant Expert powered by an external knowledge engine. (**a**) The main screen of conversation with buttons to activate voice input and display example questions, (**b**) example questions that can be asked to the chatbot as recommendations to the user.

4. Discussions

To the best of our knowledge, the presented framework is the first FAQ-powered chatbot that is fully open-source and free-to-use as well as the first that can fully operate on the client-side, which is favorable to minimize maintenance and operation costs and to protect user data. Additionally, the framework minimizes the complexity and the effort needed to implement and maintain chatbots and allow individual developers, academic research groups, small companies, non-profits, and public offices to enrich their web platforms with natural-language communication. Furthermore, information delivery enhanced with natural-language voice interactions instead of browsing through broad and entangled websites simplify the learning process for people with cognitive and mobility impairments. Another potential advantage is that the simplified design of the Instant Expert's visual elements eliminates distractions in the learning process for people with attention-deficit/hyperactivity disorder (ADHD). Finally, Q&A mode's ability to cache the model on the client-side after initial retrieval allows its usage without requiring internet connection. Hence, the research questions that are posed in this study are successfully addressed.

Instant Expert is developed as a complementary solution that focus on the gaps in existing solutions as opposed to being yet another bot development framework. In addition to stand-alone usage in Q&A mode, Knowledge-Engine mode encourages the integration with existing conversational intelligent assistant frameworks in a complementary manner. Despite its strengths, there are several limitations associated with Instant Expert. Q&A mode requires the existence of question-and-answer pairs either on a website's FAQ page or by explicit definition. However, information that users look for on a website are often provided as textual content and attachments (e.g., PDF) which requires further developments for automatic ontology building by processing the web documents with unstructured text. Another limitation of the framework is that the voice recognition via WebSpeechAPI is not supported by all browsers, and requires internet connection. To fully accomplish RQ3, open-source and client-side web libraries (e.g., PocketSphinx.js, TensorFlow.js) with voice recognition capabilities need to be reviewed and integrated within the framework as alternative services that may function offline.

5. Conclusions

This research introduces the Instant Expert, an open-source semantic web framework to effortlessly create and integrate fully functional voice-enabled smart assistants (i.e., chatbots) into any web platform with as little as a single line of HTML code. It provides a complete solution with its user interface and functionalities, communication protocols, and knowledge generation components powered by heuristic algorithms and deep learning models. It uses a variety of advanced web technologies including Web Components with Shadow DOM and CSS Containment to provide an isolated, robust, and efficient solution, and frees the developers from the complications of integrating third-party components into existing web platforms.

For future work, there are many paths to advance the proposed vision. The parsing component of the FAQ mode can be improved to (1) recursively parse question pages with links and (2) to group questions by their focus to recommend solutions to the user in cases where their question cannot directly be answered. Another potential improvement to the FAQ mode would be to allow multiple custom question variations to better model an intent so that the recall can be maximized while the knowledge base scales. The algorithm for the heuristic FAQ parsing can be further improved with using custom HTML tags or classes for question items as well as to use a machine learning model trained to recognize FAQ patterns on web pages. For the Knowledge-Engine Mode, integration modules can be developed to effortlessly connect third-party services for natural-language processing and inference. The Instant Expert's scope and capabilities can be expanded into virtual reality (e.g., A-Frame), and augmented reality (i.e., HoloLens and Magic Leap) applications. The framework can be used to scrape multiple FAQs from a variety of sources to create a smart assistant for a selected subject in any domain. Finally, web technologies, such

as WebRTC, can be used to create virtual chat rooms on the browser to allow remote participant to interact with each other while being able to invoke the chatbot for real-time information retrieval.

Author Contributions: Conceptualization, Y.S. and I.D.; methodology, Y.S.; software, Y.S.; validation, Y.S.; formal analysis, Y.S.; investigation, Y.S. and I.D.; resources, Y.S. and I.D.; data curation, Y.S.; writing—original draft preparation, Y.S.; writing—review and editing, I.D.; visualization, Y.S.; supervision, I.D.; project administration, Y.S. and I.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The framework is open-source and free to use. The repository can be accessed on GitHub (https://github.com/uihilab/instant-expert, accessed on 11 October 2021).

Acknowledgments: This project is based upon work supported by the Iowa Flood Center and the University of Iowa.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sermet, Y.; Villanueva, P.; Sit, M.A.; Demir, I. Crowdsourced approaches for stage measurements at ungauged locations using smartphones. *Hydrol. Sci. J.* **2020**, *65*, 813–822. [CrossRef]
2. Wu, X.; Zhu, X.; Wu, G.Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 97–107.
3. Sit, M.; Sermet, Y.; Demir, I. Optimized watershed delineation library for server-side and client-side web applications. *Open Geospat. Data Softw. Stand.* **2019**, *4*, 8. [CrossRef]
4. Demir, I.; Yildirim, E.; Sermet, Y.; Sit, M.A. FLOODSS: Iowa flood information system as a generalized flood cyberinfrastructure. *Int. J. River Basin Manag.* **2018**, *16*, 393–400. [CrossRef]
5. Peppard, J.; Ward, J. *The Strategic Management of Information Systems: Building a Digital Strategy*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
6. Carson, A.; Windsor, M.; Hill, H.; Haigh, T.; Wall, N.; Smith, J.; Olsen, R.; Bathke, D.; Demir, I.; Muste, M. Serious gaming for participatory planning of multi-hazard mitigation. *Int. J. River Basin Manag.* **2018**, *16*, 379–391. [CrossRef]
7. Alberts, I. Challenges of information system use by knowledge workers: The email productivity paradox. *Proc. Am. Soc. Inf. Sci. Technol.* **2013**, *50*, 1–10. [CrossRef]
8. Sermet, M.Y.; Demir, I.; Kucuksari, S. Overhead power line sag monitoring through augmented reality. In Proceedings of the 2018 North American Power Symposium (NAPS), Fargo, ND, USA, 9–11 September 2018; pp. 1–5.
9. Xu, H.; Windsor, M.; Muste, M.; Demir, I. A web-based decision support system for collaborative mitigation of multiple water-related hazards using serious gaming. *J. Environ. Manag.* **2020**, *255*, 109887. [CrossRef] [PubMed]
10. Sermet, Y.; Demir, I.; Muste, M. A serious gaming framework for decision support on hydrological hazards. *Sci. Total Environ.* **2020**, *728*, 138895. [CrossRef]
11. Liao, S.H. Expert system methodologies and applications—A decade review from 1995 to 2004. *Expert Syst. Appl.* **2005**, *28*, 93–103. [CrossRef]
12. Brandtzaeg, P.B.; Følstad, A. Why people use chatbots. In *International Conference on Internet Science*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 377–392.
13. Sit, M.; Demiray, B.Z.; Xiang, Z.; Ewing, G.J.; Sermet, Y.; Demir, I. A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.* **2020**, *82*, 2635–2670. [CrossRef] [PubMed]
14. IMARCGroup Intelligent Virtual Assistant Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2019–2024. 2019. Available online: https://www.researchandmarkets.com/reports/4775648/intelligent-virtual-assistant-market-global (accessed on 11 October 2021).
15. Schoemaker, P.J.; Tetlock, P.E. Building a more intelligent enterprise. *MIT Sloan Manag. Rev.* **2017**, *58*, 28.
16. Jain, M.; Kumar, P.; Kota, R.; Patel, S.N. Evaluating and informing the design of chatbots. In Proceedings of the 2018 Designing Interactive Systems Conference, Hong Kong, China, 9–13 June 2018; pp. 895–906.
17. Chung, K.; Park, R.C. Chatbot-based heathcare service with a knowledge base for cloud computing. *Clust. Comput.* **2019**, *22*, 1925–1937. [CrossRef]
18. Yildirim, E.; Demir, I. An integrated web framework for HAZUS-MH flood loss estimation analysis. *Nat. Hazards* **2019**, *99*, 275–286. [CrossRef]
19. Androutsopoulou, A.; Karacapilidis, N.; Loukis, E.; Charalabidis, Y. Transforming the communication between citizens and government through AI-guided chatbots. *Gov. Inf. Q.* **2019**, *36*, 358–367. [CrossRef]

20. Vaidyam, A.N.; Wisniewski, H.; Halamka, J.D.; Kashavan, M.S.; Torous, J.B. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Can. J. Psychiatry* **2019**, *64*, 456–464. [CrossRef]
21. USACE Virtual Assistant Technology Holds Promise for USACE. 2019. Available online: https://www.usace.army.mil/Media/News-Archive/Story-Article-View/Article/2014053/virtual-assistant-technology-holds-promise-for-usace/ (accessed on 11 October 2021).
22. Miner, A.S.; Laranjo, L.; Kocaballi, A.B. Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit. Med.* **2020**, *3*, 1–4. [CrossRef]
23. Sohrabi, C.; Alsafi, Z.; O'Neill, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Iosifidis, C.; Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int. J. Surg.* **2020**, *76*, 71–76. [CrossRef]
24. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [CrossRef]
25. Hwang, S.; Kim, J. Toward a Chatbot for Financial Sustainability. *Sustainability* **2021**, *13*, 3173. [CrossRef]
26. Sermet, Y.; Demir, I. An intelligent system on knowledge generation and communication about flooding. *Environ. Model. Softw.* **2018**, *108*, 51–60. [CrossRef]
27. Kolenik, T.; Gams, M. Intelligent Cognitive Assistants for Attitude and Behavior Change Support in Mental Health: State-of-the-Art Technical Review. *Electronics* **2021**, *10*, 1250. [CrossRef]
28. Lee, K.; Jo, J.; Kim, J.; Kang, Y. Can Chatbots Help Reduce the Workload of Administrative Officers?-Implementing and Deploying FAQ Chatbot Service in a University. In *International Conference on Human-Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 348–354.
29. Abdellatif, A.; Costa, D.; Badran, K.; Abdalkareem, R.; Shihab, E. Challenges in chatbot development: A study of stack overflow posts. In Proceedings of the 17th International Conference on Mining Software Repositories, Seoul, Korea, 29–30 June 2020; pp. 174–185.
30. Daniel, G.; Cabot, J.; Deruelle, L.; Derras, M. Xatkit: A multimodal low-code chatbot development framework. *IEEE Access* **2020**, *8*, 15332–15346. [CrossRef]
31. Singh, A.; Ramasubramanian, K.; Shivam, S. Introduction to Microsoft Bot, RASA, and Google Dialogflow. In *Building an Enterprise Chatbot*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 281–302.
32. Radziwill, N.M.; Benton, M.C. Evaluating quality of chatbots and intelligent conversational agents. *arXiv* **2017**, arXiv:1704.04579.
33. Schmidt, B.; Borrison, R.; Cohen, A.; Dix, M.; Gärtler, M.; Hollender, M.; Klöpper, B.; Maczey, S.; Siddharthan, S. Industrial Virtual Assistants: Challenges and Opportunities. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 8–12 October 2018; pp. 794–801.
34. MindCommerce Virtual Personal Assistants (VPA) and Smart Speaker Market: Artificial Intelligence Enabled Smart Advisers, Intelligent Agents, and VPA Devices 2019–2024. 2019. Available online: https://mindcommerce.com/reports/virtual-personal-assistant-market/ (accessed on 11 October 2021).
35. Adam, M.; Wessel, M.; Benlian, A. AI-based chatbots in customer service and their effects on user compliance. *Electron. Mark.* **2020**, *9*, 204. [CrossRef]
36. Sermet, Y.; Demir, I. Virtual and augmented reality applications for environmental science education and training. In *New Perspectives on Virtual and Augmented Reality: Finding New Ways to Teach in a Transformed Learning Environment*; Routledge: Abingdon, UK, 2020.
37. Oh, J.; Ahn, W.H.; Kim, T. Web app restructuring based on shadow DOMs to improve maintainability. In Proceedings of the 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 118–122.
38. De Ryck, P.; Nikiforakis, N.; Desmet, L.; Piessens, F.; Joosen, W. Protected web components: Hiding sensitive information in the shadows. *IT Prof.* **2015**, *17*, 36–43. [CrossRef]
39. Atkins, T.; Rivoal, F. CSS Containment Module Level 1. 2018. Available online: https://www.w3.org/TR/css-contain-1/ (accessed on 11 October 2021).
40. Damani, S.; Narahari, K.N.; Chatterjee, A.; Gupta, M.; Agrawal, P. Optimized Transformer Models for FAQ Answering. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 235–248.
41. Jijkoun, V.; de Rijke, M. Retrieving answers from frequently asked questions pages on the web. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, Bremen, Germany, 31 October–5 November 2005; pp. 76–83.
42. Farouk, M. Measuring Text Similarity Based on Structure and Word Embedding. *Cogn. Syst. Res.* **2020**, *63*, 1–10. [CrossRef]
43. Cer, D.; Yang, Y.; Kong, S.Y.; Hua, N.; Limtiaco, N.; John, R.S.; Constant, N.; Guajardo-Cespedes, M.; Yuan, S.; Tar, C.; et al. Universal sentence encoder. *arXiv* **2018**, arXiv:1803.11175.
44. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
45. Sermet, Y.; Demir, I. Towards an information centric flood ontology for information management and communication. *Earth Sci. Inform.* **2019**, *12*, 541–551. [CrossRef]

46. CNBC. Microsoft is Launching a 'Plasmabot' to Encourage People Who Recovered from the Virus to Donate Their Plasma as a Possible Treatment. Available online: https://www.cnbc.com/2020/04/18/microsoft-plasmabot-encourages-covid-19-survivors-to-donate-plasma.html (accessed on 11 October 2021).
47. Espinoza, J.; Crown, K.; Kulkarni, O. A Guide to Chatbots for COVID-19 Screening at Pediatric Health Care Facilities. *JMIR Public Health Surveill.* **2020**, *6*, e18808. [CrossRef]
48. Judson, T.J.; Odisho, A.Y.; Young, J.J.; Bigazzi, O.; Steuer, D.; Gonzales, R.; Neinstein, A.B. Case Report: Implementation of a Digital Chatbot to Screen Health System Employees during the COVID-19 Pandemic. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1450–1455. [CrossRef] [PubMed]
49. Martin, A.; Nateqi, J.; Gruarin, S.; Munsch, N.; Abdarahmane, I.; Knapp, B. An artificial intelligence-based first-line defence against COVID-19: Digitally screening citizens for risks via a chatbot. *bioRxiv* **2020**. [CrossRef] [PubMed]
50. Sharma, M.; Yadav, K.; Yadav, N.; Ferdinand, K.C. Zika virus pandemic—Analysis of Facebook as a social media health information platform. *Am. J. Infect. Control* **2017**, *45*, 301–302. [CrossRef]
51. Vergadia, P. How Can Chatbots Help during Global Pandemic (COVID-19)? Available online: https://medium.com/google-cloud/how-can-chatbots-help-during-global-pandemic-covid-19-9c1a4428d8c2 (accessed on 11 October 2021).
52. SimilarWeb. Coronavirus Data, Insights, and Trends. Available online: https://www.similarweb.com/coronavirus/ (accessed on 11 October 2021).
53. Centers for Disease Control and Prevention (CDC). Coronavirus (COVID-19) Frequently Asked Questions. 2021. Available online: https://www.cdc.gov/coronavirus/2019-ncov/faq.html (accessed on 11 October 2021).
54. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
55. He, H.; Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
56. Ross, C. I Asked Eight Chatbots Whether I Had COVID-19. The Answers Ranged from 'Low' Risk to 'Start Home Isolation'. Available online: https://www.statnews.com/2020/03/23/coronavirus-i-asked-eight-chatbots-whether-i-had-covid-19/ (accessed on 11 October 2021).
57. Microsoft. Project Answer Search. Available online: https://labs.cognitive.microsoft.com/en-us/project-answer-search (accessed on 11 October 2021).

big data and cognitive computing

MDPI

Article

DASentimental: Detecting Depression, Anxiety, and Stress in Texts via Emotional Recall, Cognitive Networks, and Machine Learning

Asra Fatima [1], Ying Li [2], Thomas Trenholm Hills [3] and Massimo Stella [1,*]

[1] CogNosco Lab, Department of Computer Science, University of Exeter, Exeter EX4 4PY, UK; af614@exeter.ac.uk
[2] Center for Adaptive Rationality, Max Planck Institute for Human Development, 14195 Berlin, Germany; li@mpib-berlin.mpg.de
[3] Department of Psychology, University of Warwick, Coventry CV4 7AL, UK; T.T.Hills@warwick.ac.uk
* Correspondence: m.stella@exeter.ac.uk

Abstract: Most current affect scales and sentiment analysis on written text focus on quantifying valence/sentiment, the primary dimension of emotion. Distinguishing broader, more complex negative emotions of similar valence is key to evaluating mental health. We propose a semi-supervised machine learning model, DASentimental, to extract depression, anxiety, and stress from written text. We trained DASentimental to identify how $N = 200$ sequences of recalled emotional words correlate with recallers' depression, anxiety, and stress from the Depression Anxiety Stress Scale (DASS-21). Using cognitive network science, we modeled every recall list as a bag-of-words (BOW) vector and as a walk over a network representation of semantic memory—in this case, free associations. This weights BOW entries according to their centrality (degree) in semantic memory and informs recalls using semantic network distances, thus embedding recalls in a cognitive representation. This embedding translated into state-of-the-art, cross-validated predictions for depression ($R = 0.7$), anxiety ($R = 0.44$), and stress ($R = 0.52$), equivalent to previous results employing additional human data. Powered by a multilayer perceptron neural network, DASentimental opens the door to probing the semantic organizations of emotional distress. We found that semantic distances between recalls (i.e., walk coverage), was key for estimating depression levels but redundant for anxiety and stress levels. Semantic distances from "fear" boosted anxiety predictions but were redundant when the "sad–happy" dyad was considered. We applied DASentimental to a clinical dataset of 142 suicide notes and found that the predicted depression and anxiety levels (high/low) corresponded to differences in valence and arousal as expected from a circumplex model of affect. We discuss key directions for future research enabled by artificial intelligence detecting stress, anxiety, and depression in texts.

Keywords: cognitive network science; text analysis; natural language processing; artificial intelligence; emotional recall; cognitive data; AI

Citation: Fatima, A.; Li, Y.; Hills, T.T.; Stella, M. DASentimental: Detecting Depression, Anxiety, and Stress in Texts via Emotional Recall, Cognitive Networks, and Machine Learning. *Big Data Cogn. Comput.* **2021**, *5*, 77. https://doi.org/10.3390/bdcc5040077

Academic Editor: Tzung-Pei Hong

Received: 26 October 2021
Accepted: 6 December 2021
Published: 13 December 2021

1. Introduction

Depression, anxiety, and stress are the three negative emotions most likely to be associated with psychopathological consequences [1–3]. According to the Depression Anxiety Stress Scale (DASS-21) [4], depression is associated with profound dissatisfaction, hopelessness, abnormal evaluation of life, self-deprecation, lack of interest, and inertia. Similarly, anxiety is associated with hyperventilation, palpitation, nausea, and physical trembling. Stress is associated with difficulty relaxing, agitation, and impatience (cf. [1,5,6]).

While these emotions are discrete and highly complex [3,6], they vary along a primary and culturally universal dimension of valence: perceived pleasantness [2,7,8]. Due to its universality, valence is frequently used in self-report affect scales [9]. Valence, which can be easily quantified along a continuous scale, explains the largest variance when compared

to other proposed emotional dimensions such as arousal [7]. Unfortunately, valence is often the only output of affect scales. This is potentially problematic for measuring depression, anxiety, and stress, which are more complex [10]: These three distinct types of psychological distress are similar in valence but differ widely along other dimensions, such as arousal [11]. In fact, depression, anxiety, and stress are difficult to distinguish using valence alone [3,12,13]. This underlines the need for richer mappings between emotional dimensions and depression, anxiety, and stress (DAS) levels.

In the present study, we examine a new approach for detecting DAS levels by exploring how they are related to the emotions people recall when asked to report how they felt recently [8]. Specifically, we developed a machine learning approach (DASentimental) to extract more comprehensive information about reported emotions from a recall-based affect scale, the Emotional Recall Task (ERT) [8]. By using DASentimental to extract information from recently experienced emotions, we show how DASentimental can be used to make inferences about DAS levels that extend beyond valence and, subsequently, allow users to investigate natural language more generally.

Literature Review: Cognitive Data Science, Mental Well-Being and Issues of Affect Scales

Mental well-being is a psychological state in which individuals are able to cope with negative stimuli and emotional states [3,6,14,15]. Assessing the connection between emotions and mental well-being is a key yet relatively unexplored dimension in cognitive data science, the branch of cognitive science investigating human psychology and mental processes under the lens of quantitative data models [6,8,16–23].

Depression, anxiety, and stress can impair mental well-being, and consequences can be as extreme as suicide [5,24]. Early quantitative recognition of distress signals that might affect mental health is crucial to providing support and boosting wellness. Recognition starts with diving deep into an individual's mindset and understanding their emotions. Psychological research [8,15,22,25,26] has found that people's mental states relate to how they communicate; written and spoken language can thus reveal psychological states. A person's emotional state can be anticipated through their communication [22,27]—identifying a quantitative coexistence and correlation among emotional words used by an individual can unveil crucial insights into their emotional state [28,29]. However, using only valence (also called "sentiment" in computer science [30]) to assess DAS levels is likely to be insufficient. Capturing how people reveal various forms of emotional distress in their natural language is therefore an important and open research area.

Outputs of affect scales typically include scores that quantify emotional valence. For example, the Positive Affect and Negative Affect Scale (PANAS; [13]), arguably the most popular self-report affect scale, asks people to evaluate their emotional experience against a predetermined emotion checklist that contains 10 positive words and 10 negative words (e.g., "to what extent did you feel irritated over the past month?"). By summing up the responses, the PANAS provides two scores: one for positive affect and one for negative affect. The PANAS essentially splits the emotions into two groups based on valence, and consequently ignores the within-group difference in valence. That means, for example, that emotions in the negative affect list such as "guilty" and "scared" are treated as though they had the same emotional impact.

Understanding mental well-being could be enhanced by both investigating richer sets of emotions—including everything a person might remember about their recent emotional experience—and examining the sequence of those emotional states. A precedent for this approach was recently set with the publication of the ERT [8], which asks participants to produce 10 emotions that describe their feelings. The sequences of words produced in the ERT represents a potential wealth of information for adapting machine learning to sentiment analysis. The idiosyncratic features of these individual words may contain information beyond valence. Indeed, arousal is often included as a primary predictor in addition to valence, for example, in the two-dimensional circumplex model of emotions [11]. Yet the ERT is likely to contain other dimensions as well. For example, anger and fear

are both highly negative and highly arousing, but they refer to different experiences and prepare people for different sets of behaviors—anger may trigger aggression whereas fear may trigger freezing or fleeing. The order in which words are recalled in the ERT may also offer useful information by indicating the availability of different emotions and therefore potentially signalling information about emotional importance [31,32]. For example, earlier-recalled words are likely to provide more information on well-being than are later-recalled words. Finally, the ERT may contain information on emotional granularity, a psychological construct referring to an individual's ability to discriminate between different emotions [14]. For example, a person with high emotional granularity would tend to use more distinct words (e.g., "anxious" instead of "bad"). People with higher emotional granularity reported better well-being and were less prone to mental illness, probably because a sophisticated understanding of one's negative emotions fosters better coping strategies [14]. Crucially, people with low emotional granularity were found to be more likely to focus on valence and use "happy" and "sad" to cover the entire spectrum of positive and negative emotions [2]. All these patterns and strategies represent the building blocks of DASentimental.

2. Research Aims

This work focuses on sequences of emotional words, whose ordering and semantic meaning contain features that are assumed to be predictive of depression, anxiety, and stress. Having defined these psychological constructs along the psychometric scale represented by the DASS-21, the current work aims to reconstruct the model between emotional word sequences and DAS levels through machine learning. We adopted a semi-supervised learning approach mainly composed of two stages. First, we trained a machine learning regression model using cognitive data from the ERT [8]. Through cross-validation and feature selection, we enriched word sequences with a cognitive network representation [15,33] of semantic memory. We show that semantic prominence in the recall task as captured by network degree can boost the performance of the regression task. Second, having selected the best-performing model, we applied it to identifying emotional sequences in text, providing estimations for the DAS levels of narrative/emotional corpora—in this case, suicide notes [24,28].

3. Methods

This section outlines the datasets and methodological approaches adopted in this manuscript.

3.1. Datasets: Emotional Recall Data, Free Associations, Suicide Notes, and Valence–Arousal Norms

Four datasets were used to train and test DASentimental: the ERT dataset [8], the Small World of Words free association dataset in English [33], the corpus of genuine suicide notes curated by Schoene and Dethlefs [28], and valence–arousal norms in the Valence-Arousal-Dominance (VAD) Lexicon [10].

The ERT dataset is a collection of emotional recalls provided by 200 individuals and matched against psychometric scales such as the DASS-21 [4]. During the recall task, each participant was asked to produce a list of 10 words expressing the emotions they had felt in the last month. Participants were also asked to assess items on psychometric scales, thus providing data in the form of word lists/recalls (e.g., (anger, hope, sadness, disgust, boredom, elation, relief, stress, anxiety, happiness) and psychometric scores (e.g., depression/anxiety/stress levels between 0 and 20). The completely anonymous dataset makes it possible to map the sequences of emotional words recalled by individuals against their mental well-being, achieved here through a machine learning approach.

The Small World of Words [33] project is an international research project aimed at mapping human semantic memory [19] through free associations: conceptual associations where one word elicits the recall of others [33]. Cognitive networks made of free associations between concepts have been successfully used to predict a wide variety of cognitive phenomena, including language processing tasks [33], creativity levels [18,20], early word

learning [34,35], picture naming in clinical populations [36], and semantic relatedness tasks [37,38]. Furthermore, because they have no specific syntactic or semantic constraints, free associations capture a wide variety of associations encoded in the human mind [39]. We therefore used free associations to model the structure of semantic memory from which the ERT recalls were selected. This modeling approach posits that all individuals, independent of their well-being, possess a common structure of conceptual associations. Although preliminary evidence shows that semantic memory might be influenced by external factors such as distress [40] or personality traits [41], we had to adopt this point as a necessary modeling simplification in absence of free association norms across clinical populations. Our approach also posits that the connectivity of emotional words is not uniform; rather, there are more (and less) well connected concepts. This postulation, combined with the adoption of a network structure, operationalized the task of identifying how semantically related emotional words are in terms of network distance (the length of the shortest path connecting any two node [35]). The finding that network distance in free associations outmatched semantic latent analysis when modeling semantic relatedness norms [37,38] supports our approach.

The corpus of suicide notes is a collection of 142 suicide notes by people who ended their lives [28]. The dataset was curated and analyzed for the first time by Schoene and Dethlefs [28], who used it to devise a supervised learning approach to automatic detection of suicide ideation. The notes were collected from various sources, including newspaper articles and other existing corpora. All notes were anonymized by removing any links to a person or place or any other identifying information. Already investigated in previous studies under the lens of sentiment analysis [28], cognitive network science [24], and recurrent neural networks [29], this dataset was a clinical case study to which DASentimental was applied after having been trained on word sequences from the ERT data.

The valence–arousal norms used here indicate how pleasant/unpleasant (valence) and how exciting/inhibiting (arousal) words are when identified in isolation within a psychology mega-study [10]. This dataset included valence and arousal norms for over 20,000 English words. We used it to validate, through the circumplex model of affect [11], results based on DASentimental and text analysis.

3.2. Machine Learning Regression Analysis

Our DASentimental approach aims at extracting depression, anxiety, and stress levels from a given text through semi-supervised learning. DASentimental is a regression model, trained on features extracted from emotional recalls (ERT data) and obtained from a network representation of semantic memory (free association data). The model was validated against psychometric scores from the DAS scale [8]. Using cross-validation and feature importance analysis [42], we selected a best-performing model to detect depression, anxiety, and stress levels in previously unseen texts.

All in all, the pipeline implemented in this work can be divided into four main subtasks, performed sequentially:

1. Data cleaning and vectorial representation of regressor (features) and response (DAS levels) variables;
2. Training, cross-validation, and selection of the best-performing regression model for estimating DAS levels from ERT data;
3. Estimating the DAS levels of suicide notes by parsing the sequences of emotional words in each letter;
4. Validating the labelling predicted by DASentimental through independent affective norms [10].

3.3. Data Cleaning and Vectorial Representation of Regressor Variables

Our regression task builds a mapping between depression (anxiety, stress) scores $\{Y\}_i$ and features extracted from sequences of emotional words, $\{X\}_{\{i,j\}}$. Each sequence contains exactly 10 words produced by a participant in the ERT (e.g., X_i = *{anger, hope,*

sadness, disgust, boredom, elation, relief, stress, anxiety, happiness} and thus X_{ij} = *anger*). The ERT dataset features 200 recalls X_i, each produced by an individual and reported relative to their estimated levels of depression, anxiety, and stress as measured on the DASS-21 (cf. [8]). One training instance was performed for each of the three independent variables (depression, anxiety, and stress levels). All instances used 200 recalls in a fourfold cross-validation. Independent variables ranged from 0 (e.g., absence of stress) to 20 (e.g., high levels of stress). The ERT dataset contained on average one in three individuals suffering from abnormal levels (for reference see: https://www.psytoolkit.org/survey-library/depression-anxiety-stress-dass.html accessed on 8 November 2021) of depression, anxiety, or stress, indicating that the dataset contains both normal and abnormal levels of distressing emotional states and can be used for further analysis.

Like other approaches in natural language processing [20,23,42], we adopted a vectorial representation, transforming the 10-dimensional vectors X_i into N-dimensional vectors B_i, where the first $K < N$ entries B_{ik} (for $1 \leq k \leq K$) count the occurrence (1, 2, 3, ...) or absence (0) of a word in the original recall list X_i. In this way, K counts the first entries in the vectorial representation of recalls featuring the absence or presence of specific words. Consequently, K is also the number of unique words present across all 200 recalls in the ERT data. The remaining entries B_{ik} (for $K + 1 \leq k \leq N$) are relative to additional features extracted from recall lists, i.e., network distances obtained from the cognitive network of free associations. Hence, N is the sum of K and the number of distance-based features extracted from recalls. The representation of word lists as binary vectors of word occurrences, known as Bag-of-Words (BOW) [43], is one of the simplest and most commonly used numerical representations of texts in natural language processing. The representation of word lists as features extracted from a network structure, known as network embedding, has been used in cognitive network science for predicting creativity levels from animal category tasks [20].

BOW representations can be noisy due to different word forms indicating the same lexical item and thus the same semantic/emotional content of a list (e.g., "depressed" and "depression"). Noise in textual data can be reduced by regularizing the text—that is, recasting different words to the same lemma or form. We cast different forms to their noun counterparts through the WordNet lemmatization function implemented in the Natural Language Toolkit and available in Python. This data cleaning reduced the overall set of unique words from 526 to K = 355 nouns, thus reducing the dimensionality and sparsity of our vector representations.

3.4. Embedding Recall Data in Cognitive Networks of Free Associations

A crucial limitation of the BOW representation is that emotional words have the same weight for the regression analysis regardless of whether they were recalled first or last. This is in contrast with a rich literature on recall from semantic memory [19,21], which indicates that in producing a list of items from a given category, the elements recalled first are generally more semantically relevant to the category itself. Therefore a better refinement would be to weight word entries in BOW according to their position in recalls. For instance, the occurrence of "sad" in the first position in recall i (i.e., X_{i1} = sad) would receive a higher weight w_1 than if it occurred later. The different weights $\{w\}_j$ could be tailored so that initial words in a recall are more important in estimating DAS levels.

Rather than using arbitrary weightings, we adopted a cognitive network science approach [17]. Emotional words do not come from an unstructured system; they are the outcome of a search in human memory [21]. We modeled this memory as a network of free associations in order to embed words in a network structure and measure their relevance in memory through network metrics (cf. [19,21,44]). This approach allowed us to compute the network centrality of all words in a given position j and estimate weight w_j as the average of such scores.

Our first step was to transform continuous free association data from [33] into a network where nodes represented words and links represented memory recall patterns

(e.g., word A reminding at least two individuals of word B). Analogously to other network approaches [34,35,37,38] and due to the asymmetry in gathering cues and targets [39], we considered links as being undirected. This procedure led to a representation of semantic memory as a fully connected network $\mathcal{N}$ containing 34,298 concepts and 328,936 links. On this representation we then computed semantic relevance through one local metric (degree) and one global metric (closeness) that had been adopted in previous cognitive inquiries [18,34,36]. Degree captures the numbers of free associations providing access to a given concept, whereas closeness centrality identifies how far on average a node is from its connected neighbors (cf. [17]). We checked that all $K = 355$ unique words from the ERT data were present in $\mathcal{N}$, then computed degrees and closeness centralities of all words occurring in a given position $j \in \{1, 2, \ldots, 10\}$. We report the results in Figure 1.

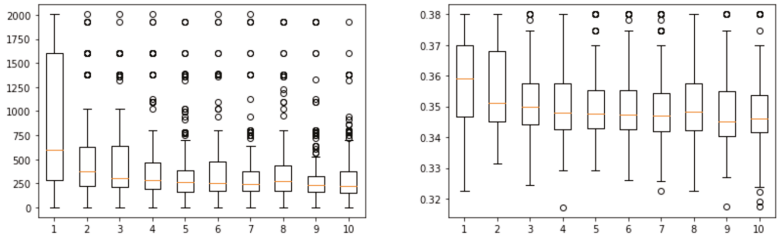


Figure 1. Box plots of degree (**left**) and closeness centrality (**right**) of words in the ERT dataset for each position of recall, $j \in \{1, 2, \ldots, 10\}$.

Although there are several outliers, Figure 1 confirms previous findings [8] about the ERT data following memory recall patterns, with words in the first positions being more semantically prominent than subsequent ones. Since degree and closeness centrality did not seem to display qualitatively different behaviors, we used the median values m_j for degree (see Figure 1, left) as weights $w_j = m_j$, normalized so that $\sum w_j = 1$. These weights were used to multiply/weight the respective entries of the BOW representation, B_{ik} (for $1 \leq k \leq K$), so as to obtain a weighted BOW representation of recalls depending on both the ERT data and the network representation $\mathcal{N}$ of semantic memory.

This procedure constitutes a first semantic embedding of ERT data in a cognitive network. We also performed a second semantic embedding of ERT recalls by considering them as walks over the structure of semantic memory. Like the approach in [20], we considered a list of recalled words as a network path [17], $X_{i1}, X_{i2}, \ldots, X_{i10}$, visiting nodes over the network structure and moving along shortest network paths [38]. This second embedding made it possible to attribute a novel set of distance-based features to each recall.

In particular, we focused on the following distance-based features:

1. The coverage C_i performed over the whole walk X_i [36]—that is, the total number of free associations traversed when navigating $\mathcal{N}$ across a shortest path from node X_{i1} to X_{i2}, then from X_{i2} to X_{i3}, and so on. This coverage equals the sum or the total of all the network distances between adjacent words in a given recall;
2. The graph distance entropy [45] E_i of the whole walk X_i, computed as the Shannon entropy for the occurrences of paths of any length within X_i;
3. The total network distance D_i between all nodes in a walk/recall X_i and the target word "depression". Similarly, we also considered S_i (A_i) as the sum of distances between recalled words and "stress" ("anxiety");
4. The total network distance H_i between all nodes in a walk/recall X_i and the target emotional state "happy". Similarly, we also considered SS_i (F_i) as the sum of distances between recalled words and the target emotion "sad" ("fear").

We considered these metrics following previous investigations of semantic memory, affect, and personality traits. Coverage on cognitive networks has been found to be an important metric for predicting creativity levels with recall tasks [16,20]. Higher coverage

and graph distance entropy can indicate that sets of responses are more scattered across the structure of semantic memory or that they oscillate between positive and negative emotional states, with potential repercussions over reduced emotion regulation and increased DAS levels [8]. Since shortest network distance on free association networks was found to predict semantic similarity [37,38], we selected semantic distances between recalls and target clinical states to capture the relatedness of responses to DAS levels. The selection of "happy" and "sad" followed previous results from the circumplex model of affect [11], a model mapping emotional states according to the dimensions of pleasantness and arousal. In the circumplex model, "happy" and "sad" are opposite emotional states and their relatedness to recalls can provide additional information for detecting the presence or absence of states such as DAS. We also included "fear", as it is a common symptom of DAS disorders [8].

The validation of these distances as useful features for discriminating between DAS levels is presented in the Results section.

3.5. Machine Learning Approaches

After having built weighted and unweighted BOW representations of ERT data, enriched with distance-based measured, we tested them within three commonly adopted machine learning regressors.

We tested the following algorithms [42,46]: (i) decision tree, (ii) multilayer perceptron (MLP), and (iii) recurrent neural network (Long–Short Term Memory [LSTM]). Decision trees can predict target values by learning decision rules on how to partition a dataset according to its features in order to reduce the total error between predictions and estimates (cf. [23]). The MLP is inspired by biological neural networks [47] and consists of multiple computing units organized in input, hidden, and output layers. Each unit takes a linear combination of features and produces an output according to an activation function. Combinations are fixed according to weights that are updated over time so as to minimize the error between the final and target inputs. This procedure, known as backpropagation, travels backwards on the neural network [46]. LSTM networks feature feed-forward and back-forward loops that affect hidden layers recurrently over training, a procedure known as deep learning. Additionally, LSTMs feature specific nodes that remember outputs over arbitrary time intervals; this can enhance training by reducing the occurrence of vanishing gradients and getting stuck in local minima.

3.6. Model Training

We trained decision trees with *scikit-learn* in Python [42]. We used a maximum tree depth of 8 to reduce overfitting and applied a squared error optimization metric for identifying tree nodes. For MLPs we selected an architecture using two hidden layers with 25 neurons each. A dropout rate of 20% between weight updates in the second hidden layer was fixed to reduce overfitting. The number of layers and neurons were fixed after fine-tuning over multiple iterations using the whole dataset of 200 data points and a fourfold cross-validation. A rectified linear activation function was selected to keep the output positive at each layer, as is the case with DAS scores. Training was performed via backpropagation [46]. For the LSTM architecture, we used two hidden layers, each featuring four cells and a dropout rate of 20% to reduce overfitting. A batch gradient descent algorithm was used to train the LSTM network [46].

Training was performed by splitting the dataset of 200 ERT recalls into training (75%) and test sets (25%), according to a fourfold cross-validation. In the regression task of estimating DAS levels from the test set after training, we measured performance in terms of mean squared error (MSE) loss and Pearson's correlation R. Vectors of features underwent an L2 regularization to further reduce the impact of large dimensionality and sparseness during regression. Performance with different sets of features was recorded so we could apply the best-performing model to text analysis.

3.7. Application of DASentimental to Text

Texts are sequences of words, albeit in more articulated forms than sequential recalls from semantic memory. Nonetheless, word co-occurrences in texts are not independent from semantic memory structure itself—in fact, a growing body of literature in distribution semantics adopts co-occurrences for predicting free association norms [48]. We adopted an analogous approach and used the best-performing model from the ERT data to estimate DAS levels in texts based on their sequences of emotional words. DASentimental can thus be considered a semi-supervised learning approach, trained on psychologically validated recalls and applied to previously unseen sequences of emotional words in texts.

To enhance overlap between the emotional jargon of text and the lexicon of $K = 355$ unique emotional words in the ERT dataset, we implemented a text parser in spaCy, identifying tokens in texts and mapping them to semantically related items in the ERT lexicon. This semantic similarity was obtained as a cosine similarity between pre-trained word2vec embeddings; it was therefore independent from free association distances.

As seen in Algorithm 1, for every non-stopword [49] of every sentence in a text, the parser identifies nouns, verbs, and adjectives and maps them onto the most similar concept (if any) present in the ERT/DASentimental lexicon. This procedure can skip stopwords and enhance the attribution of different word forms and tenses to their corresponding base form from the ERT dataset, which possesses less linguistic variability than does text due to its recall-from-memory structure. This mapping helped ensure that DASentimental does not miss different forms of words or synonyms from texts, and therefore ultimately enhances the quality of the regression analysis. Checking for item similarity in network neighborhoods drastically reduced computation times, thereby making DASentimental more scalable for volumes of texts larger than the 142 notes used in this first study.

Algorithm 1: Semantic parser identifying emotional words from text that can be mapped onto the emotional lexicon of DASentimental.

```
   Input: Text from Suicide Note
   Output: Vector Representation of emotional content, selected words
 1 for each sentence in suicide note do
 2   for each word in sentence do
 3     if word is negative:
 4       isNeg = True
 5     if word not in stopwords and word.pos in ['NOUN','ADJ','ADV','VERB']
 6       if isNeg True:
 7         find similar words to the current word antonym in ERT words
 8         if max similarity >= 0.5:
 9           Add most similar word to selected words and update vector
10           isNeg=False
11       else:
12         find similar words to the current word in ERT words
13         if max similarity >= 0.5:
14           Add most similar word to selected words and update vector
15   end for
16 end for
```

3.8. Handling Negations in Texts

Using spaCy also meant we could track negation in a sentence and, for any emotional words in the same sentence, substitute in their antonym to be checked (instead of the non-negated word). A similar approach was adopted in studies with cognitive networks [50]. For instance, in the sentence "I am not happy", the word "happy" was not directly checked for similarity against the ERT lexicon. Instead, the antonym of "happy" ("sad") was found using spaCy and its similarity was checked instead. Handling negations is a key aspect of

processing texts. Since more elaborate forms of meaning negations are present in language, this can be considered a first, simple approach to accounting for semantic negations.

3.9. Psycholinguistic Validation of DASentimental for Text Analysis

In this first study we used suicide letters as a clinical corpus investigated in previous works [28,50] and featuring narratives produced by individuals affected by pathological levels of distress. Unfortunately but understandably, the corpus did not feature annotations expressing the levels of depression, anxiety, and stress felt by the authors of the letters.

We also performed emotional profiling [51] over the same set of suicide notes, this time relying on another psycholinguistic set of affective norms, the NRC VAD Lexicon [10]. Analogously to the emotional profiling implemented in [30] to extract key states from textual data, we used the NRC VAD Lexicon to provide valence and arousal scores to lemmatized words occurring in suicide notes. We also applied DASentimental to all suicide notes and plotted the resulting distributions of depression (anxiety, stress) scores. A qualitative analysis of the distributions highlighted tipping points, which were used for partitioning the data into letters indicating high and low levels of estimated depression (anxiety, stress). Tipping points were selected instead of medians because most notes elicited no estimated DAS levels, thus producing imbalanced partitions. The tipping points were 6 for depression, 2 for anxiety, and 4 for stress. As reported in Figure 2 (bottom), above these tipping points the distributions exhibited cut-offs or abrupt changes.

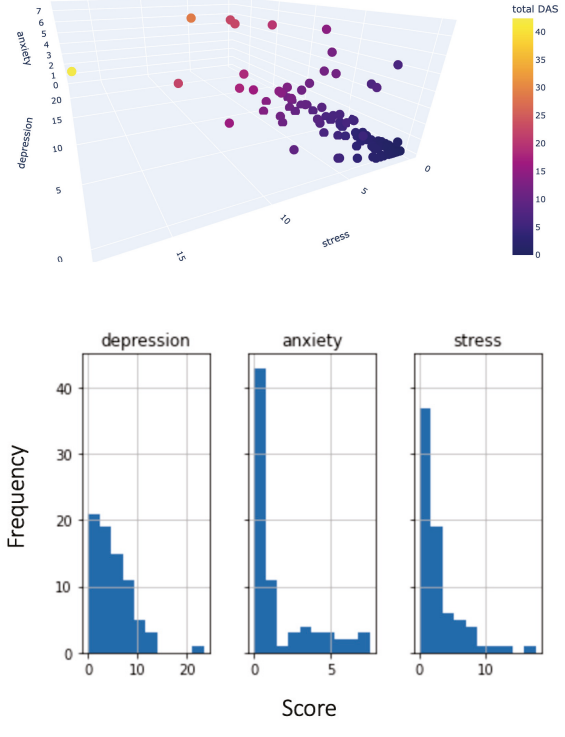


Figure 2. Top: 3D visualization of depression, anxiety, and stress in suicide notes as estimated by DASentimental. **Bottom**: Histograms of DAS levels per pathological construct.

We then compared the median valence and arousal of words occurring in high and low partitions of the suicide note corpus. Our exploration was guided by the circumplex

model of affect [11], which maps "depression" as a state with negative valence and low arousal, "anxiety" as a state with negative valence and high arousal, and "stress" as a state in between "anxiety" and "depression". For every partition, we expected letters tagged as "high" by DASentimental to feature more extreme language.

4. Results

This section reports on the main results of the manuscript. First, we quantify semantic distances and their relationships with DAS levels. Second, we outline a comparison of different learning methods. Third, within the overall best-performing machine learning model we compare performance of the binary and weighted BOW representations of recalls, using only the embedding coming from network centrality. We then provide key results about several models using different combinations of network distances, further enriching the ERT data with features coming from network navigation of semantic memory (see Methods). We conclude by applying the best-performing model to the analysis of suicide letters and presenting the results of the psycholinguistic validation of DASentimental estimations.

4.1. Semantic Distances Reflect Patterns of Depression, Anxiety, and Stress

Semantic distances, in the network representation of semantic memory, correlated with DAS levels. In other words, the emotional words produced by individuals tended to be closer to or further from targets such as "depression" and "anxiety" (see Methods) according to the DAS levels recorded via the psychometric scale. We found a Pearson's correlation coefficient R between depression levels and total semantic distance between recalls and "depression" equal to -0.341 ($N = 200$, $p < 0.0001$). This means that people affected by higher levels of depression tended to recall and produce emotional words that were semantically closer and thus more related to [37,38] the concept "depression" in semantic memory. We found similar patterns for "anxiety" and anxiety levels (-0.218, $N = 200$, $p = 0.002$) and for "stress" and stress levels (-0.357, $N = 200$, $p < 0.0001$). These findings constitute quantitative evidence that semantic distance from these target concepts can be useful features for predicting DAS levels. For the happy/sad emotional dimension, we found that people with higher depression/anxiety/stress levels tended to produce concepts closer to "sad" ($R < -0.209$, $N = 200$, $p < 0.001$). Only people affected by lower depression levels tended to recall items closer to "happy" ($R = 0.162$, $N = 200$, $p = 0.02$). No statistically significant correlations were found for anxiety and stress levels. These results indicate that the sad/happy dimension might be particularly relevant for estimating depression levels. At a significance level of 0.05, no other correlations were found. Because there might be additional correlations between different features that are exploitable by machine learning, we will further test the relevance of distance over machine learning regression within the trained models.

The relevance of semantic distances for predicting DAS levels can also be visualized by partitioning the ERT dataset into individuals with higher-than-median or lower-than-median depression (anxiety, stress) levels. Figure 3 (bottom) shows that people with lower levels tended to produce distributions of network distances differing in their medians (Kruskal–Wallis test, $N_H + N_L = 200$, $p < 0.001$). Individuals with higher levels of depression (anxiety, stress) tended to recall items more semantically prominent to "depression" ("anxiety", "stress"), which further validates our correlation analysis and our adoption of network distance as features for regression.

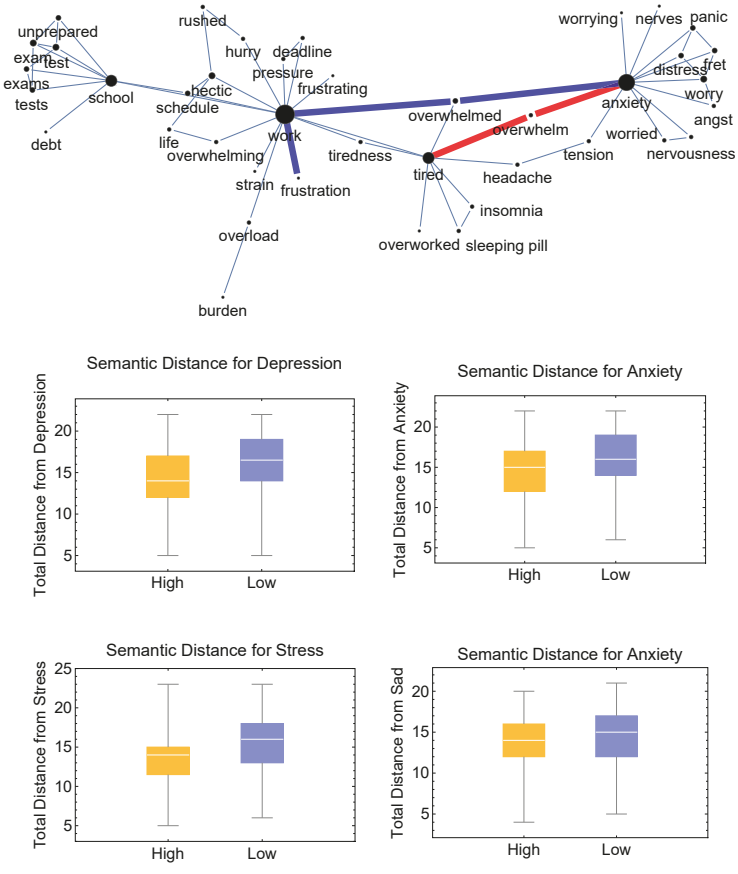


Figure 3. Top: Toy representation of semantic memory as a network of free associations. Semantic network distances from recalls (e.g., "tired", "frustration") to "anxiety" are highlighted. This visualization illustrates that cognitive networks provide structure to conceptual organization in the mental lexicon and enable measurements such as semantic relatedness in terms of shortest paths/network distance. **Bottom**: Total network distances between recalls and individual concepts ("anxiety", "depression", "stress", and "sad") between people with high and low levels of DAS.

4.2. Performance of Different Machine Learning Algorithms

As reported in the Methods section, we trained ERT data through three machine learning algorithms: (i) decision trees, (ii) LSTM recurrent neural networks and (iii) MLP. Independent of using the binary or embedded BOW representations of ERT recall and tuning hyperparameters, neither decision trees nor LSTM networks learned from the dataset. This might be due to the relatively small size of the sample (200 recalls). It must be also noted that decision trees try to split the data based on a single feature value, whereas in the current case DAS levels might depend on the co-existence of emotional words. For instance, the sequence "love, broken" might portray love in a painful context, creating a codependence of features that would be difficult for decision trees to account for. The MLP did learn relationships from the data; its performance is outlined in the next section in terms of MSE loss and R^2 between estimations and validation values of DAS.

4.3. Embedding BOW in Semantic Memory Significantly Boosts Regression Performance

Table 1 reports the average performance of the MLP regressor over binary and weighted representations of word recalls from the ERT. In our approach, weights come from the median centralities of words in the network representation of semantic memory enabled by free associations [19,38,39] (see Methods).

The binary BOW representation of recalls achieved nontrivial regression results (R^2 higher than 0) for the estimation of depression levels, but not for anxiety or stress estimation. Enriching the same vector representation with the weights from cognitive network science drastically boosted performance, with R^2 ranging between 0.15 and 0.40 (i.e., R between 0.38 and 0.63). These results indicate that the first items recalled from semantic memory possess more information about the DAS levels of a given individual. This indicates that there is additional structure within the ERT data that we capitalize on by using the weighted representation.

Table 1. Average losses and R^2 estimators for the binary and weighted versions of Bag-of-Words (BOW) representations of ERT recalls. Weights were fixed according to the median centralities of words in each position of the ERT data (see Methods). Error margins are computed over 10 iterations and indicate standard deviations. MSE: mean squared error.

DAS Constructs	Binary BOW		Cognitive Weighted BOW	
	MSE Loss	R^2	MSE Loss	R^2
Depression	30.7 ± 0.1	0.19 ± 0.01	22.0 ± 0.1	0.40 ± 0.02
Anxiety	16.2 ± 0.1	0.03 ± 0.01	14.5 ± 0.1	0.15 ± 0.02
Stress	27.6 ± 0.1	0.03 ± 0.01	19.3 ± 0.1	0.26 ± 0.01

4.4. Comparison of Model Performance Based on Cognitive Network Features

Table 2 reports model performance when different network distances were plugged in together with the weighted BOW.

Considering all semantic distances was beneficial for boosting regression results, reducing MSE loss, and enhancing R^2 levels—up to 0.49 for depression levels ($R = 0.7$), 0.20 for anxiety levels ($R = 0.44$), and 0.27 for stress levels ($R = 0.52$). The artificial intelligence trained here correlates as strongly as the ERT metric introduced by Li and colleagues [8], which relied on additional valence data attributed to emotions by participants and was powered by traditional statistical methods not based on machine learning. Since additional valence data is not available in texts, we did not use it to train DASentimental; nevertheless, we achieved analogous prediction correlations to Li and colleagues. We therefore considered DASentimental to be a state-of-the-art tool for assessing depression, anxiety, and stress levels from emotional recall data, and proceeded to the next step: using it for text analysis.

Table 2 also identifies the relevance of different distance-based features for predicting DAS levels. We found that the addition of coverage was beneficial for boosting prediction performance compared to the weighted BOW only; this might be due to nonlinear effects that cannot be captured by the previous regression analysis. Adding happy/sad distances to coverage worsened prediction results for "depression" and, in general, produced a lower boost than did adding distances from depression/anxiety/stress to the unweighted BOW representation. Adding all these distances introduced feature correlations that the MLP exploited to achieve higher performance.

Furthermore, Table 2 reports crucial results for exploring how "fear" relates to the estimation of DAS levels. The semantic distances from fear produced a boost in predicting anxiety, indicating that fear is an important emotion for predicting anxiety levels. No boost was recovered for other DAS constructs. However, because these distances are correlated with others, the two models using "fear" and all other concepts performed as well as the simpler model without "fear". We therefore selected the model based on the weighted

BOW representation plus coverage/entropy and all other distances except from "fear" as the final model of DASentimental.

Table 2. Average losses and R^2 estimators for models employing different features within the same neural network architecture. All models include weighted representations of word and include (in order of appearance in table): (i) all network distances/conceptual entries (from "anxiety", "depression", "stress", "sad|", "happy", and "fear", together with total emotional coverage (Cover.)), (ii) only distances from "depression", "stress", and "anxiety" with coverage and graph distance entropy (Entr.), (iii) only distances from "happy" and "sad" with coverage and entropy, (iv) only coverage and entropy, (v) all other distances, coverage, and entropy without distances from "fear", and (vi) only distances from "fear". Error margins are computed over 10 iterations and indicate standard deviations. BOW: bag-of-Words, MSE: mean squared error.

Cognitive-Network Embedded BOW with:	Construct	MSE Loss	R^2
All Conceptual Distances + Cover. + Entr.	Depression	18.6 ± 0.4	0.49 ± 0.01
	Anxiety	14.3 ± 0.3	0.20 ± 0.02
	Stress	19.3 ± 0.3	0.27 ± 0.01
Only Distances from Depression/Anxiety/Stress + Cover. + Entr.	Depression	19.5 ± 0.5	0.46 ± 0.01
	Anxiety	14.4 ± 0.3	0.17 ± 0.02
	Stress	19.0 ± 0.4	0.28 ± 0.02
Only Distances from Happy/Sad + Cover. + Ent.	Depression	20.6 ± 0.5	0.43 ± 0.01
	Anxiety	14.9 ± 0.2	0.15 ± 0.01
	Stress	19.3 ± 0.4	0.27 ± 0.01
Only Cover. + Entr.	Depression	19.5 ± 0.6	0.45 ± 0.01
	Anxiety	14.7 ± 0.2	0.15 ± 0.01
	Stress	19.2 ± 0.4	0.27 ± 0.02
Cover. + Entr. + All Distances except from Fear	Depression	18.5 ± 0.3	0.49 ± 0.01
	Anxiety	13.9 ± 0.3	0.23 ± 0.02
	Stress	18.9 ± 0.5	0.28 ± 0.01
Distance from Fear only	Anxiety	14.6 ± 0.2	0.16 ± 0.01

4.5. Analysis of Suicide Notes

According to the World Health Organization (See: https://www.who.int/news-room/fact-sheets/detail/suicide; last accessed 5 October 2021), every year more than 700,000 people commit suicide. There are usually multiple reasons behind a person's decision to take their life, and one reason can lead to others: Minor incidents may accumulate over time, increasing mental distress and the appearance of depression, anxiety, and/or stress. Eventually, these may become too overwhelming to endure, and the mental pressure may trigger the decision to end one's own life [5].

Most people who commit suicide do not leave behind a note; those who do provide vital information. Written by individuals who have reached the limit of emotional distress, suicide notes are first-hand evidence of the vulnerable mindset of emotionally distraught individuals [28,29]. Analyzing these notes can offer important insights into the mental states of their authors.

To gather such insights, we applied the best-performing version of DASentimental (with weighted BOW and semantic distances) to the corpus of genuine suicide notes curated by Schoene and Deathlefs [28] and investigated in other recent studies [24,29]. This application constitutes the second part of our semi-supervised approach to text analysis, in which DASentimental predicts DAS levels of non-annotated text from its semantically enriched sequences of emotional words (see Methods).

Results are reported in Figure 2. We registered strong positive correlations between estimated DAS levels (Pearson's coefficients, $R_{DA} = 0.35$, $p < 0.0001$; $R_{DS} = 0.50$, $p < 0.0001$; $R_{AS} = 0.59$, $p < 0.0001$). These correlations indicate that suicide notes tended to feature similar levels of distress coming from depression, anxiety, and stress, although with different intensities and frequencies, as evident from the qualitative analysis of distributions in Figure 2 (bottom).

Using valence and arousal of words expressed in suicide letters, we performed an additional validation of the results of DAS through the circumplex model of affect [11] (see Methods). By partitioning the notes according to high/low levels of depression (anxiety, stress), we compared the valence (and arousal) of all words mentioned in suicide letters from each partition. At a significance level of 0.05, suicide notes marked by DASentimental as showing higher depression levels were found to contain a lower median valence than notes marked as showing lower levels of depression (Kruskal–Wallis test, $KS = 6.889$, $p = 0.009$). Analogously, suicide notes marked by DASentimental as showing higher anxiety levels contained a higher median arousal than notes marked as showing lower anxiety levels (Kruskal–Wallis test, $KS = 3.2014$, $p = 0.007$). No differences for stress were found.

Letters with lower depression levels contained more positive jargon, including mentions of loved ones and "relief" at ending the pain and starting a new chapter. Some notes even included emotionless instructions about relatives and assets to be taken care of. Letters with higher depression levels more frequently mentioned jargon relative to "pain" and "boredom"—this imbalance in frequency is captured by the difference in median valence noted above. In the circumplex model, depression lives in a space with more negative valence than neutral/emotionless language; thus the statistically significant difference in valence indicates that DASentimental is able to identify the negative dimension associated with depression.

A similar pattern was found for anxiety, with letters marked as "high anxiety" by DASentimental featuring more anxious jargon relative to pain and suffering. In the circumplex model, anxiety lives in a space with higher arousal and alarm than neutral/emotionless language; therefore, the difference in median arousal between high- and low-anxiety letters indicates that DASentimental is able to identify the alarming and arousing dimensions associated with anxiety.

The absence of differences for stress might indicate that DASentimental is not powerful enough to detect these differences; if this is the case, further research and larger datasets are required. Nonetheless, the signals of enhanced negativity and alarm detected by DASentimental lay the foundation for detecting stress, anxiety, and depression in texts via emotional recall data.

5. Discussion

In this study, we trained a neural network, DASentimental, to predict depression, anxiety, and stress using sequences of emotional words embedded in a cognitive network representation of semantic memory. DASentimental achieved cross-validated predictions for depression ($R = 0.7$), anxiety ($R = 0.44$), and stress ($R = 0.52$) in line with previous approaches using additional valence data [8]. This state-of-the-art performance suggests that even without explicitly encoding valence ratings for each word, DASentimental is able to achieve good explanatory power. This success stems from the cognitive embedding of recalled concepts—that is, concepts at shorter/longer network distance from key ideas such as "depression", "sad", and "happy" that are commonly used to describe emotional distress [4,7].

Our findings suggest the importance of considering network distances between concepts in semantic memory in investigations of emotional distress. This insight provides further support to studies showing how network distances and connectivity can predict other cognitive phenomena, including creativity levels [16,18,20], semantic distance [37,38], and word production in clinical populations [36].

We noticed a significant boost in performance (+210% in R^2 on average) when embedding BOW representations of recall lists (see Methods) in a cognitive network of free associations [33]. A nearly tenfold boost was observed for predicting stress and anxiety, which are considerably complex distress constructs [14,40]. Our results underline the need to tie together artificial intelligence/text-mining [43] and cognitive network science [15] to achieve cutting-edge predictors in next-generation cognitive computing.

We applied DASentimental to a collection of suicide notes as a case study. Most suicide notes in the corpus [28] indicated low levels of depression, anxiety, and stress. This suggests that despite the decision to terminate their own life, the writers of suicide notes tried to avoid overwhelming their last messages with negative emotions; this is compatible with previous studies [24]. One observation gained from a close reading of suicide notes is that many writers expressed their love and gratitude to their significant others, and used euphemisms when referencing the act of suicide (e.g., "I can't take this anymore"). Therefore, although a reader might interpret a typical suicide note as being filled with sorrow, their perception is influenced by the knowledge that the writer eventually killed themself.

Limitations and Future Research

A key limitation of DASentimental is that it cannot account for how context shifts and forges meaning and perceptions in language [43]. Furthermore, DASentimental cannot capture how the writers actually felt before, during, or after writing those last letters. Instead, DASentimental quantifies the emotions that are explicitly expressed by the authors, since it is trained on ERT data which includes expressions of emotions without context. Future research might better detect contextual knowledge through natural language processing, which has been successfully used on contextual features such as medical reports [52] and speech organization [27] to detect the risk of psychosis in clinical populations. Alternatively, community detection in feature-rich networks could provide information about the different meanings and contextual interpretations provided to concepts in cognitive networks (e.g., the different meanings of "star" [53]). Last but not least, contextual features might be detected through meso-scale network metrics such as entanglement, which was recently shown to efficiently identify nodes critical for information diffusion in a variety of techno-social networks [54].

Subjective valence ratings such as those adopted in the original ERT study [8] are not available in texts; DASentimental therefore uses cognitive network distances [45] to target words instead. Despite this difference, DASentimental's performance is on par with the work of Li and colleagues [8]. This has two implications: First, a stronger model might use distances and valences when focusing only on fluency tasks. Second, for text analysis and even cognitive social media mining [55], DASentimental's machine learning pipeline could be used to detect any kind of target emotion (e.g., "surprise" or "love"). Furthermore, DASentimenal could be used on future datasets relating DAS levels with ERT data and demographics to explore how age, gender, or physical health can influence depression, anxiety, and stress detection.

In this first study we relied on machine learning to spot relationships between emotional recall and depression, anxiety, and stress levels. However, other statistical approaches might be employed in the future—the most promising is LASSO for semi-supervised regression [56].

As a future research direction, DASentimental could be used to investigate the cultural evolution of emotions. Emotions and their expressions are shaped by culture and learned in social contexts [57,58] and media movements [49]. What people can feel and express depends on their surrounding social norms. Previous studies have shown that large historical corpora can be used to make quantitative inferences on the rise and fall of national happiness [58]. Similarly, DASentimental could be applied to track the change of explicit expression of depression, anxiety, and stress over history, quantified through the emotions of modern individuals. This would highlight changes in norms towards

emotional expression and historical events (e.g., "pandemic"), thus complementing other recent approaches in cognitive network science [9,30,59–61] and sentiment/emotional profiling [51,55,62,63] by bringing to the table a quantitative, automatic quantification of depression, anxiety, and stress in texts.

6. Conclusions

This work combines cognitive network science and artificial intelligence to introduce a semi-supervised machine learning model, DASentimental, that extracts depression, anxiety, and stress from written text. Cognitive data and networked representations of semantic memory powered our approach. We trained the model to spot how sequences of recalled emotion words by $N = 200$ individuals correlated with their responses to the DASS-21. Weighting responses according to their network centrality (degree) and measuring recalls as network walks significantly boosted regression results up to state-of-the-art approaches (cf. [8]). This quantitative approach not only makes it possible to assess emotional distress in texts (e.g., detecting the depression level expressed in a letter), but also provides a quantitative framework for testing how semantic distances of emotions correlate with distress and mental well-being. For these reasons, DASentimental will appeal to clinical researchers interested in measuring distress levels in texts in ways that are interpretable under the framework of semantic network distances (cf. [16,20,37,38]).

Author Contributions: Conceptualization, A.F. and M.S.; methodology, all authors; software, A.F.; validation, all authors; formal analysis, A.F. and M.S.; investigation, A.F. and M.S.; writing—original draft, review and editing, all authors; visualization, A.F.; supervision, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki over datasets already released in previous ethically approved studies.

Informed Consent Statement: Since no novel data was generated from this study, no informed consent was required.

Data Availability Statement: We prepared a Python package containing the best model of DASentimental implemented and described in this paper. The package is available on GitHub: https://github.com/asrafaiz7/DASentimental (accessed on 21 November 2021). If using DASentimental for estimating anxiety, stress and depression levels in texts, please cite this manuscript.

Acknowledgments: We acknowledge Deborah Ain for kindly providing comments and help with manuscript editing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lovibond, P.F.; Lovibond, S.H. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Ther.* **1995**, *33*, 335–343. [CrossRef]
2. Russell, J.A.; Barrett, L.F. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J. Personal. Soc. Psychol.* **1999**, *76*, 805. [CrossRef]
3. O'Driscoll, C.; Buckman, J.E.; Fried, E.I.; Saunders, R.; Cohen, Z.D.; Ambler, G.; DeRubeis, R.J.; Gilbody, S.; Hollon, S.D.; Kendrick, T.; et al. The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: Analysis of data from six randomised control trials. *BMC Med.* **2021**, *19*, 1–14. [CrossRef] [PubMed]
4. Akin, A.; Çetın, B. The Depression Anxiety and Stress Scale (DASS): The study of Validity and Reliability. *Educ. Sci. Theory Pract.* **2007**, *7*, 260–268.
5. Conejero, I.; Olié, E.; Calati, R.; Ducasse, D.; Courtet, P. Psychological pain, depression, and suicide: Recent evidences and future directions. *Curr. Psychiatry Rep.* **2018**, *20*, 1–9. [CrossRef]
6. Abend, R.; Bajaj, M.A.; Coppersmith, D.D.; Kircanski, K.; Haller, S.P.; Cardinale, E.M.; Salum, G.A.; Wiers, R.W.; Salemink, E.; Pettit, J.W.; et al. A computational network perspective on pediatric anxiety symptoms. *Psychol. Med.* **2021**, *51*, 1752–1762. [CrossRef]
7. Barrett, L.F. Valence is a basic building block of emotional life. *J. Res. Personal.* **2006**, *40*, 35–55. [CrossRef]

8. Li, Y.; Masitah, A.; Hills, T.T. The Emotional Recall Task: Juxtaposing recall and recognition-based affect scales. *J. Exp. Psychol. Learn. Mem. Cogn.* **2020**, *46*, 1782–1794. [CrossRef]
9. Montefinese, M.; Ambrosini, E.; Angrilli, A. Online search trends and word-related emotional response during COVID-19 lockdown in Italy: A cross-sectional online study. *PeerJ* **2021**, *9*, e11858. [CrossRef]
10. Mohammad, S. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 174–184.
11. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. [CrossRef]
12. Tellegen, A. *Structures of Mood and Personality and Their Relevance to Assessing Anxiety, with an Emphasis on Self-Report*; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 1985.
13. Watson, D.; Clark, L.A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1988**, *54*, 1063. [CrossRef]
14. Tugade, M.M.; Fredrickson, B.L.; Feldman Barrett, L. Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *J. Personal.* **2004**, *72*, 1161–1190. [CrossRef]
15. Kenett, Y.N.; Faust, M. Clinical cognitive networks: A graph theory approach. In *Network Science in Cognitive Psychology*; Routledge: London, UK, 2019; pp. 136–165.
16. Beaty, R.E.; Zeitlen, D.C.; Baker, B.S.; Kenett, Y.N. Forward Flow and Creative Thought: Assessing Associative Cognition and its Role in Divergent Thinking. *Think. Ski. Creat.* **2021**, *41*, 100859. [CrossRef]
17. Siew, C.S.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, *2019*, 2108423 . [CrossRef]
18. Kenett, Y.N.; Levy, O.; Kenett, D.Y.; Stanley, H.E.; Faust, M.; Havlin, S. Flexibility of thought in high creative individuals represented by percolation analysis. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 867–872. [CrossRef]
19. Kumar, A.A. Semantic memory: A review of methods, models, and current challenges. *Psychon. Bull. Rev.* **2021**, *28*, 40–80. [CrossRef]
20. Stella, M.; Kenett, Y.N. Viability in multiplex lexical networks and machine learning characterizes human creativity. *Big Data Cogn. Comput.* **2019**, *3*, 45. [CrossRef]
21. Hills, T.T.; Jones, M.N.; Todd, P.M. Optimal foraging in semantic memory. *Psychol. Rev.* **2012**, *119*, 431. [CrossRef]
22. Golino, H.F.; Epskamp, S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE* **2017**, *12*, e0174035.
23. Tohalino, J.A.; Quispe, L.V.; Amancio, D.R. Analyzing the relationship between text features and grants productivity. *Scientometrics* **2021**, *126*, 4255–4275. [CrossRef]
24. Teixeira, A.S.; Talaga, S.; Swanson, T.J.; Stella, M. Revealing semantic and emotional structure of suicide notes with cognitive network science. *arXiv* **2020**, arXiv:2007.12053.
25. Zemla, J.C.; Cao, K.; Mueller, K.D.; Austerweil, J.L. SNAFU: The semantic network and fluency utility. *Behav. Res. Methods* **2020**, *52*, 1681–1699. [CrossRef]
26. Morgan, S.E.; Diederen, K.; Vertes, P.E.; Ip, S.H.; Wang, B.; Thompson, B.; Demjaha, A.; De Micheli, A.; Oliver, D.; Liakata, M.; et al. Assessing psychosis risk using quantitative markers of disorganised speech. *medRxiv* **2021**. [CrossRef]
27. Morgan, S.; Diederen, K.; Vértes, P.; Ip, S.; Wang, B.; Thompson, B.; Demjaha, A.; De Micheli, A.; Oliver, D.; Liakata, M.; et al. Natural Language Processing markers in first episode psychosis and people at clinical high-risk. *Transl. Psychiatry* **2021**. [CrossRef]
28. Schoene, A.M.; Dethlefs, N. Automatic identification of suicide notes from linguistic and sentiment features. In Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Berlin, Germany, 11 August 2016; pp. 128–133.
29. Schoene, A.M.; Turner, A.; De Mel, G.R.; Dethlefs, N. Hierarchical Multiscale Recurrent Neural Networks for Detecting Suicide Notes. *IEEE Trans. Affect. Comput.* **2021**. [CrossRef]
30. Stella, M.; Restocchi, V.; De Deyne, S. # lockdown: Network-enhanced emotional profiling in the time of COVID-19. *Big Data Cogn. Comput.* **2020**, *4*, 14.
31. Pachur, T.; Hertwig, R.; Steinmann, F. How do people judge risks: Availability heuristic, affect heuristic, or both? *J. Exp. Psychol. Appl.* **2012**, *18*, 314. [CrossRef] [PubMed]
32. Tversky, A.; Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cogn. Psychol.* **1973**, *5*, 207–232. [CrossRef]
33. De Deyne, S.; Navarro, D.J.; Perfors, A.; Brysbaert, M.; Storms, G. The "Small World of Words" English word association norms for over 12,000 cue words. *Behav. Res. Methods* **2019**, *51*, 987–1006. [CrossRef] [PubMed]
34. Hills, T.T.; Maouene, M.; Maouene, J.; Sheya, A.; Smith, L. Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychol. Sci.* **2009**, *20*, 729–739. [CrossRef]
35. Stella, M.; Beckage, N.M.; Brede, M. Multiplex lexical networks reveal patterns in early word acquisition in children. *Sci. Rep.* **2017**, *7*, 1–10.

36. Castro, N.; Stella, M.; Siew, C.S. Quantifying the interplay of semantics and phonology during failures of word retrieval by people with aphasia using a multiplex lexical network. *Cogn. Sci.* **2020**, *44*, e12881. [CrossRef]
37. Kumar, A.A.; Balota, D.A.; Steyvers, M. Distant connectivity and multiple-step priming in large-scale semantic networks. *J. Exp. Psychol. Learn. Mem. Cogn.* **2020**, *46*, 2261. [CrossRef]
38. Kenett, Y.N.; Levi, E.; Anaki, D.; Faust, M. The semantic distance task: Quantifying semantic distance with semantic network path length. *J. Exp. Psychol. Learn. Mem. Cogn.* **2017**, *43*, 1470. [CrossRef]
39. De Deyne, S.; Navarro, D.J.; Storms, G. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behav. Res. Methods* **2013**, *45*, 480–498. [CrossRef]
40. Smith, A.M.; Hughes, G.I.; Davis, F.C.; Thomas, A.K. Acute stress enhances general-knowledge semantic memory. *Horm. Behav.* **2019**, *109*, 38–43. [CrossRef]
41. Kenett, Y.; Baker, B.; Hills, T.; Hart, Y.; Beaty, R. Creative Foraging: Examining Relations Between Foraging Styles, Semantic Memory Structure, and Creative Thinking. In Proceedings of the Annual Meeting of the Cognitive Science Society, Vienna, Austria, 25–29 July 2021; Volume 43.
42. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
43. Hassani, H.; Beneki, C.; Unger, S.; Mazinani, M.T.; Yeganegi, M.R. Text mining in big data analytics. *Big Data Cogn. Comput.* **2020**, *4*, 1. [CrossRef]
44. Hills, T.T.; Kenett, Y.N. Networks of the Mind: How Can Network Science Elucidate Our Understanding of Cognition? *Top. Cogn. Sci.* **2021**. [CrossRef]
45. Stella, M.; De Domenico, M. Distance entropy cartography characterises centrality in complex networks. *Entropy* **2018**, *20*, 268. [CrossRef]
46. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2020.
47. Gardner, M.W.; Dorling, S. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [CrossRef]
48. Vankrunkelsven, H.; Verheyen, S.; De Deyne, S.; Storms, G. Predicting lexical norms using a word association corpus. In Proceedings of the 37th Annual Conference of the Cognitive Science Society, Pasadena, CA, USA, 22–25 July 2015; pp. 2463–2468.
49. Amancio, D.R.; Oliveira, O.N., Jr.; da Fontoura Costa, L. Identification of literary movements using complex networks to represent texts. *New J. Phys.* **2012**, *14*, 043029. [CrossRef]
50. Stella, M. Text-mining forma mentis networks reconstruct public perception of the STEM gender gap in social media. *PeerJ Comput. Sci.* **2020**, *6*, e295. [CrossRef]
51. Mohammad, S.M. Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. In *Emotion Measurement*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 323–379.
52. Irving, J.; Patel, R.; Oliver, D.; Colling, C.; Pritchard, M.; Broadbent, M.; Baldwin, H.; Stahl, D.; Stewart, R.; Fusar-Poli, P. Using natural language processing on electronic health records to enhance detection and prediction of psychosis risk. *Schizophr. Bull.* **2021**, *47*, 405–414. [CrossRef]
53. Citraro, S.; Rossetti, G. Identifying and exploiting homogeneous communities in labeled networks. *Appl. Netw. Sci.* **2020**, *5*, 1–20. [CrossRef]
54. Ghavasieh, A.; Stella, M.; Biamonte, J.; De Domenico, M. Unraveling the effects of multiscale network entanglement on empirical systems. *Commun. Phys.* **2021**, *4*, 1–10. [CrossRef]
55. Stella, M. Cognitive network science for understanding online social cognitions: A brief review. *Top. Cogn. Sci.* **2021**. [CrossRef]
56. Jung, A.; Vesselinova, N. Analysis of network lasso for semi-supervised regression. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, Naha, Japan, 16–18 April 2019; pp. 380–387.
57. Morini, V.; Pollacci, L.; Rossetti, G. Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study. *Appl. Sci.* **2021**, *11*, 5390. [CrossRef]
58. Hills, T.T.; Proto, E.; Sgroi, D.; Seresinhe, C.I. Historical analysis of national subjective wellbeing using millions of digitized books. *Nat. Hum. Behav.* **2015**, *3*, 1271–1275. [CrossRef]
59. Simon, F.M.; Camargo, C.Q. Autopsy of a metaphor: The origins, use and blind spots of the 'infodemic'. *New Media Soc.* **2021**. [CrossRef]
60. Li, Y.; Luan, S.; Li, Y.; Hertwig, R. Changing emotions in the COVID-19 pandemic: A four-wave longitudinal study in the United States and China. *Soc. Sci. Med.* **2021**, *285*, 114222. [CrossRef] [PubMed]
61. Cinelli, M.; Quattrociocchi, W.; Galeazzi, A.; Valensise, C.M.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The COVID-19 social media infodemic. *Sci. Rep.* **2020**, *10*, 1–10.
62. Semeraro, A.; Vilella, S.; Ruffo, G. PyPlutchik: Visualising and comparing emotion-annotated corpora. *PLoS ONE* **2021**. [CrossRef] [PubMed]
63. Radicioni, T.; Squartini, T.; Pavan, E.; Saracco, F. Networked partisanship and framing: A socio-semantic network analysis of the Italian debate on migration. *arXiv* **2021**, arXiv:2103.04653.

big data and cognitive computing

MDPI

Article

Extraction of the Relations among Significant Pharmacological Entities in Russian-Language Reviews of Internet Users on Medications

Alexander Sboev [1,2,*], Anton Selivanov [1], Ivan Moloshnikov [1], Roman Rybka [1], Artem Gryaznov [1], Sanna Sboeva [1] and Gleb Rylkov [1]

[1] National Research Centre "Kurchatov Institute", 123182 Moscow, Russia; Selivanov_AA@nrcki.ru (A.S.); Moloshnikov_IA@nrcki.ru (I.M.); Rybka_RB@nrcki.ru (R.R.); Gryaznov_AV@nrcki.ru (A.G.); Sboeva_SG@rrcki.ru (S.S.); Rylkov_GV@rrcki.ru (G.R.)
[2] Moscow Engineering Physics Institute, National Research Nuclear University, 115409 Moscow, Russia
* Correspondence: Sboev_AG@nrcki.ru

Abstract: Nowadays, the analysis of digital media aimed at prediction of the society's reaction to particular events and processes is a task of a great significance. Internet sources contain a large amount of meaningful information for a set of domains, such as marketing, author profiling, social situation analysis, healthcare, etc. In the case of healthcare, this information is useful for the pharmacovigilance purposes, including re-profiling of medications. The analysis of the mentioned sources requires the development of automatic natural language processing methods. These methods, in turn, require text datasets with complex annotation including information about named entities and relations between them. As the relevant literature analysis shows, there is a scarcity of datasets in the Russian language with annotated entity relations, and none have existed so far in the medical domain. This paper presents the first Russian-language textual corpus where entities have labels of different contexts within a single text, so that related entities share a common context. therefore this corpus is suitable for the task of belonging to the medical domain. Our second contribution is a method for the automated extraction of entity relations in Russian-language texts using the XLM-RoBERTa language model preliminarily trained on Russian drug review texts. A comparison with other machine learning methods is performed to estimate the efficiency of the proposed method. The method yields state-of-the-art accuracy of extracting the following relationship types: ADR–Drugname, Drugname–Diseasename, Drugname–SourceInfoDrug, Diseasename–Indication. As shown on the presented subcorpus from the Russian Drug Review Corpus, the method developed achieves a mean F1-score of 80.4% (estimated with cross-validation, averaged over the four relationship types). This result is 3.6% higher compared to the existing language model RuBERT, and 21.77% higher compared to basic ML classifiers.

Keywords: pharmacological text corpus; automatic relation extraction; natural language processing; deep learning

Citation: Sboev, A.; Selivanov, A.; Moloshnikov, I.; Rybka, R.; Gryaznov, A.; Sboeva, S.; Rylkov, G. Extraction of the Relations among Significant Pharmacological Entities in Russian-Language Reviews of Internet Users on Medications. *Big Data Cogn. Comput.* **2022**, *6*, 10. https://doi.org/10.3390/bdcc6010010

Academic Editor: Min Chen

Received: 31 October 2021
Accepted: 27 December 2021
Published: 17 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The developing ecosystem of social networks and other special Internet platforms expands the possibility of discussion of a broad set of topics in textual format. These texts often contain people's publicly available opinions on various subjects. One of the topics of special interest is Internet reviews on medications, including information about their positive and adverse effects, qualities, manufacturers, administration regime etc. Such information could be useful for comprehensive analysis for the purposes of pharmacovigilance [1] and potential medicine re-profiling.

Analysing such a large amount of information is a time-consuming task, therefore requiring methods for automated extraction of pharmacologically-meaningful data. In

turn, these methods require textual corpora with annotation of pharmacological entities and their relations.

There is a wide variety of English-language datasets in literature sources, for example Drug—Drug Interaction (DDI) and Adverse Drug Event (ADE). These corpora contain pharmaceutically relevant entities of different types as well as relationships between them. A more detailed analysis of the corpora is presented in Section 2. However, currently there is only one large domain-oriented dataset in the Russian language: Russian Drug Review Corpus of Internet User Reviews with Complex NER labeling (RDRS), which was presented by our group [2,3]. Now, we present (in Section 3.1) an extension of this corpus that includes annotation of relationships among the named entities that are most relevant for the potential studies of drug efficiency.

The automation of the process of extracting meaningful information from a review written in a natural language requires solving the following tasks: text segmentation, Named Entity Recognition (NER), Relation Extraction (RE), structuring of the extracted information, and evaluation of the results. In this paper, we focus on the relation extraction task. The formulation of the relations extraction task in natural language processing is as follows: given a text and two entities from it, determine if there is a relation of a certain type between the entities. For example, in the text "Antiviral syrup for children Orvirem—we have an allergy to it!" with the entities "Orvirem" and "allergy", the task is to determine that the allergy is mentioned as the adverse effect of the "Orvirem" medication.

The relation extraction task can be solved by two approaches: the sequential (cascade) approach of solving the named entity recognition and relation extraction tasks separately, or the combined approach of solving these tasks simultaneously (the combined approach, called "joint" or "end-to-end" in the literature). The sequential solution allows estimating the accuracy of solving each task separately, thus leading to more thorough analysis of the task complexity; therefore, the scope of our research is to analyze the relation extraction model within the sequential approach, applied to entities already extracted.

Our review (see Section 2) showed that the most promising technology that can be utilized to solve the relation extraction task is deep learning. This paper uses a model based on the XLM-RoBERTa language model, pre-trained on a huge unlabelled corpus of drug reviews. Section 3 contains the details of the model configuration and setting.

Based on this model, a set of computational experiments are performed (Section 5) on different parts of the RDRS corpus. In Section 5.1, the optimal model parameters and text representation are obtained using a part of the corpus that includes texts with ADR–Drugname relations. Section 5.2 presents evaluations on a subset of the corpus containing reviews with multiple contexts. This experiment is aimed at obtaining the state-of-the art results for the task of relation extraction for the following four relation types: ADR–Drugname, Drugname–Diseasename, Drugname–SourceInfoDrug, Diseasename–Indication. The results of the proposed model are compared with the results of the existing language model RuBERT, as well as a set of baseline methods: multinomial naive bayes classifier, linear support vector machine, and dummy classifiers.

The main contributions of our work include:

- The relation extraction method is proposed, in which the task of determining the presence of a relation is formulated using multi-context annotation: entities belonging to the same context are considered to be related. The method is based on a language model fine-tuned to classify entity pairs by the presence of relations.
- Several variations of the text representation used to present the entities under consideration to the language model are compared, and the optimal representation is shown to be the one that includes the text of target entities along with the whole review text, concatenated with special tokens;
- The method based on a language model trained on a large corpus of unlabeled Russian drug review texts and fine-tuned on an annotated corpus of Russian drug reviews is shown to be applicable to the task of determining the relations among

pharmaceutically-relevant entities of the newly-created corpus. The accuracy estimations are obtained for this task for Russian language;
- The same proposed model, pre-trained on Russian drug reviews, is shown to achieve relation extraction results comparable to the state of the art on the DDI corpus.

2. Related Works

Currently, efficient natural language processing (NLP) methods are mostly based on neural network algorithms [4,5]. There is a wide variety of text analysis tasks that neural networks can solve with high accuracy, such as part of speech tagging [6], machine translation [7], authorship attribution [8], named entity recognition [9–13], entity relation extraction [14–16] and so on. In this paper, our research is aimed at the relation extraction task for the pharmacological domain.

There are two main approaches to the extraction of the relations between entities from a text:

- cascade approach: sequential text analysis, where the tasks of named entity recognition and relation extraction are performed separately. At the first stage, named entities are extracted from the text, either by expert annotation or using a machine learning model [3,9,13,17,18]. At the second stage, the entities extracted are evaluated in terms of their possible relations [14,15,19,20]. This approach allows one to control the learning process of each model, which in turn gives the opportunity for a more thorough choice of methods and hyperparameters.
- "joint" approach, where a single model is used to solve both tasks simultaneously [16,21–23]. The most commonly used neural network topology for this model has two parts (one for each task) that learn jointly using combined loss function.

This work is focused on the separate analysis of the relation extraction method which could be used as a part of a cascade model.

The development of tools for textual data analysis depends on the annotated data necessary for tuning the algorithms and assessing their performance. There are a number of corpora of textual data in English language with a markup of pharmacologically-relevant entities and relationships. These corpora differ by the types of texts (online reviews, tweets, clinical discharges etc.) and by the types of the named entities and relationships annotated, which are specified with varying levels of detail. Some studies provided the achievable accuracy of extracting relationships between pharmacologically-relevant entities using methods developed on base of these corpora.

Among the datasets on biomedical topics with markup suitable for solving the problem of identifying relationships between named entities, it is necessary to mention the corpora of the i2b2 Competition Corps Workshop on Natural Language Processing Challenges for Clinical Records. The competition is organized by the Department of Biomedical Informatics (DBMI) at Harvard Medical School. This organization provides datasets called n2c2, that consist of full texts of clinical records in English. The data annotation is enriched at every new competition as the scope of the competition expands and changes.

The task of extracting relationships between the named entities was considered in the 2009 [24], 2010 [25] and 2018 [26] corpora from the above-mentioned competition. The best-performing models used additional sources of information, handcrafted and engineered features, which facilitated better classification of the entities and relations in the medical discharges in English. On the Drug-Drug Interaction (DDI) dataset [27], which contains excerpts about drug interactions from the DrugName and MedLine databases, a model [28] based on a BERT-type language model SciBert [29] was used to solve the task of classifying the sentences for relationships between the selected drugs. The model is, in fact, the BERT language model, additionally trained on scientific texts for better representation of the thematic. The model showed a result of 84.08% by the f1-micro metric.

On the Adverse Drug Event (ADE) dataset [30], which contains sentences from the abstracts of PubMed scientific articles with relations between medical drugs and their adverse reactions, the performance of a BERT-based model SpERT [16] was presented. The

model solves the entity extraction and relation extraction problems sequentially. To solve the problem of identifying named entities, all possible consecutive sets of words in the text (of limited length) are generated and then classified by the model according to the type of the entity. The entities obtained as the result of the classification are then filtered, forming pairs of entities for which the model determines the presence of a relationship. The model uses BERT to represent words as vectors. Word vectors are then processed by a fully-connected layer with softmax activation function and with the size equal to the number of the named entity types. This layer identifies named entities from the input word combinations. The second part of the model is a fully-connected neural network layer with the sigmoid activation function, the input of which is the concatenation of the entity vectors with the vector representation of words between the entities. Relations are determined by applying the threshold to the output activities. Such model achieves the f1-macro metric of 79.24% on the ADE dataset.

Regarding the Russian language, a fairly limited set of corpora for relation extraction tasks are publicly available. However, these corpora facilitate the apriori assessment of the accuracy in the extraction of the relationships between named entities of different types, not related to pharmaceuticals: RuSERRC [31]—80 manually annotated texts with entities from the computer science subjects (software, database, programming languages, etc.). RuREBus [32]—300 annotated texts of strategic programs of the Ministry of Economic Development of the Russian Federation, containing various relations between the entities of the following types: Social Objects, Actions, Goals, Tasks; RURED [33]—a corpus of 536 annotated texts about economics, containing the entity types of Geographic Objects, Names, Age, Currencies, etc., as well as the relationships of various types between them; Factrueval [19]—255 annotated texts with entities of type Persons, Locations and Organizations, and also relations: Ownership, Occupation, Meeting, and Deal; NEREL [34]—933 annotated documents with the markup of a large number of entities, including: Persons, Organizations, geopolitical entities, numbers, dates, time, money, age, etc., as well as relations between them.

On the RuSERC corpus (split by sentences), a BERT-based architecture, R-BERT [14], was used to obtain the result of 67% by the macro-f1 metric. On the RuRBus corpus (split by documents), the R-BERT architecture [14] was used to obtain the result of 44% by the micro-f1 metric.

The R-BERT model uses the BERT model to represent words as vectors. A vector representing a named entity is the average vector of its words. A concatenated pair of entities is presented as the input for the fully-connected layer, and the activities of the layer are the input for the softmax layer which is used to determine the type of a relation between the entities.

On the RUED dataset (split by sentences), the span-BERT architecture achieved 78% accuracy by the f1 metric (the method of aggregating f1 across different classes was not specified). On the Factrueval dataset (split by documents), that method achieved 66% accuracy on the fact extraction task (extracting relationships among multiple entities). On the NEREL dataset (split by documents), the RuBERT model achieved the f1-score value of 51% (the method of aggregating f1 across different classes was not specified).

As for the Russian-language corpora annotated to extract the relationships between pharmacologically significant entities, the only corpus of this type is the Russian Drug Review Corpus (RDRS 2800 reviews), which is considered in this paper. Therefore, the accuracy demonstrated in the works above with other types of texts is only an estimate of the possible accuracy of determining the relationships between pharmacological entities, which is an additional motivation to perform the present work.

Summarizing the above, it can be concluded that the current trend in identifying relationships between named entities is the use of models with transformer architecture pretrained on large datasets. Further in this work, we develop this approach based on the XLM-RoBERTa language model [35] using the Russian Drug Review Corpus (RDRS) [3] described in Section 3.1 and available at the Sagteam project website (https://sagteam.ru/u).

3. Materials and Methods

3.1. Datasets

This paper uses the Russian Drug Review Corpus (RDRS) [3], which contains 2800 texts of drug reviews written by Internet users. The corpus contains markup for 18 types of named entities, which can be divided into 3 groups:

- Medication—this group includes everything related to the mentions of drugs and drugs manufacturers, including: Drug name, Drug class, Drug form, Route (how to use the drug), Dosage, SourceInfoDrug (source of the consumer's information about the drug) etc.;
- Disease—this group contains entities related to the diseases or reasons for using the drug (disease name, indications or symptoms), as well as mentions of the effects achieved (NegatedADE—the drug was inefficient, Worse—some deterioration was observed, BNE-POS—the condition improved) etc.
- ADR—mentions of occurring adverse reactions.

In a subset of the corpus containing 1590 review texts, entities were marked up into "lines of meaning"—"contexts", so that each context contains entities that describe the usage of some medication by one person for the treatment of one condition. Different contexts arise in a text, in particular, when describing the use of multiple drugs in the treatment, or different effects following the use of a single drug for different conditions, or when the review describes the use of a drug by different people. In terms of the relation extraction problem, entities that occur in the same context are related, while entities from different contexts are considered unrelated.

An example of the context annotation is shown in Figure 1. The main (1st) context of the review is about the drug "orvirem" which caused an allergy. This context includes the following mentions (denoted on the figure with a number 1 above them): "antiviral" (Drugclass), "syrup" (Drugform) "orvirem" (Drugname), multiple mentions of "allergy" (ADR), "red spots" (ADR), "swelling on the face" (ADR), "1 day" (Duration). There are other contexts in the review:

- 2nd context: "allergy" (Diseasename), "red spots" (Indication), "zyrtek" (Drugname), "the situation did not improve" (NegatedADE), "it seems to have gotten even worse" (Worse).
- 3d context: "allergy" (Diseasename), "red spots" (Indication), "doctor" (SourceInfoDrug), "On her recommendation" (SourceInfoDrug), "smecta" (Drugname), "the situation did not improve" (NegatedADE), "it seems to have gotten even worse" (Worse).
- 4th context: "allergy" (Diseasename), "red spots" (Indication), "doctor" (SourceInfoDrug), "On her recommendation" (SourceInfoDrug), "suprastin" (Drugname), "the situation did not improve" (NegatedADE), "it seems to have gotten even worse" (Worse), "Injected" (Route), "The redness seems to pass" (BNE-POS), ""swelling on the face still remains" (NegatedADE).
- 5th context: "allergy" (Diseasename), "red spots" (Indication), "doctor" (SourceInfoDrug), "prednisone" (Drugname), "Injected" (Route), "The redness seems to pass" (BNE-POS), ""swelling on the face still remains" (NegatedADE).

In Tables 1–4 the quantitative characteristics of the corpus with contextual markup are presented.

Table 1. The number of texts that contain the corresponding number of contexts.

Contexts Count	1	2	3	>3
Texts Count	682	559	218	131

Table 2. Average lengths of contexts in the corpus.

	Average Mentions Count	Average Tokens Count
Main context	19.9	38.9
Other contexts	3.7	6.6

Table 3. Statistics on the part of RDRS dataset that is comprised of ADR–Drugname relations.

Number of	Train	Test
Texts	502	126
Sentences	4016	1008
Words	82,425	20,961
"ADR" type entities	1461	356
"Drugname" type entities	1416	368
Relations	3444	845
Avg. numbers of relations per text	6.9	6.7

Table 4. Statistics on the types of relations in the RDRS corpus with 908 multi-context reviews.

Relation	ADR & Drugname		Drugname & Diseasename		Drugname & SourceInfoDrug		Diseasename & Indication	
Classes	**pos.**	**neg.**	**pos.**	**neg.**	**pos.**	**neg.**	**pos.**	**neg.**
Relations count	1913	917	4277	2153	2700	1232	2588	701
Text fraction	0.273	0.204	0.634	0.514	0.598	0.457	0.416	0.148

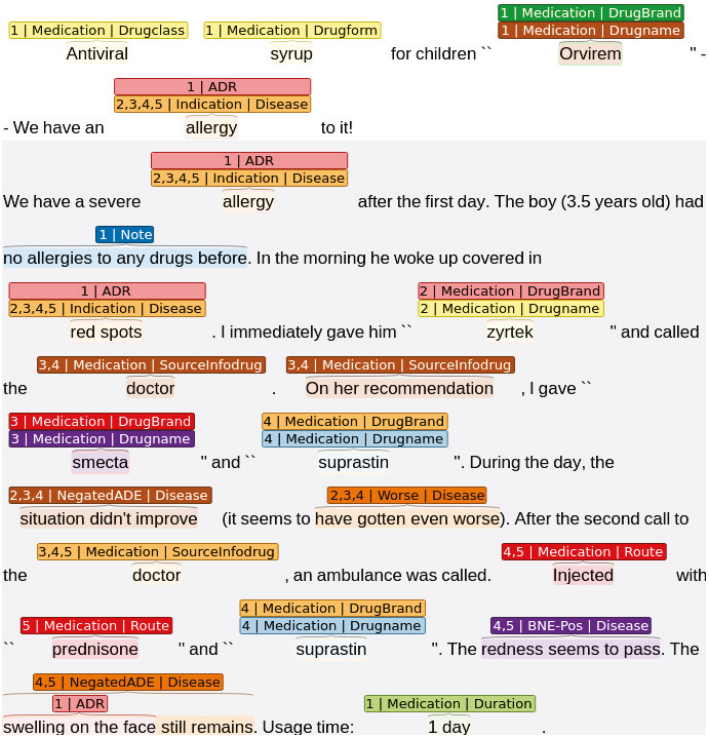


Figure 1. Example of an annotated review. The labels contain, from left to right: context number, entity type, attribute within the entity type.

In this paper, the following pairs of entities are chosen as the most interesting to analyze from the practical point of view:

- ADR–Drugname—the relationship between the drug and its side effects;
- Drugname–SourceInfodrug—the relationship between the medication and the source of information about it (e.g., "was advised at the pharmacy", "the doctor recommended it");
- Drugname–Diseasename—the relationship between the drug and the disease;
- Diseasename–Indication—the connection between the illness and its symptoms (e.g., "cough", "fever 39 degrees").

Two subsets of the original corpus have been compiled for the experiments:

1. The first one includes 628 texts containing ADR and Drugname entity pairs. The experiments on this part are aimed at selecting the most effective combinations of the input feature representations and hyper-parameters of the methods used. The texts of the RDRS corpus that contain ADR and Drugname entities were divided into training and test parts, the composition of which is presented in Table 3.
2. The second part includes texts that contain multiple contexts. The total number of such texts is 908. Statistics on the types of relationships are presented in Table 4. This corpus is used to establish the current level of accuracy in determining the relationships between pharmacologically-significant entities in Russian-language review texts.

Experiments with these subsets are described further in Section 4.

3.2. Methods

3.2.1. Deep Learning Methods

Language Models

In this work the XLM-RoBERTa-sag model [3] was used. The original XLM-RoBERTa [35] is a multilingual language model based on the transformer [36] architecture, consisting of multihead attention layers which create vector representations of the input data parts (words in case of NLP) that encode the information about their context. XLM-RoBERTa is trained on a large multilingual corpus from the CommonCrawl project that contains 2.5 TB of texts. XLM-RoBERTa-sag is a result of additional training of XLM-RoBERTa on a dataset of unlabeled internet texts about medicines (~1.65 M texts).

During the adjustment experiments, we used two versions of the model:

- XLM-RoBERTa-base-sag—12 Transformer blocks, 768 hidden neurons, 8 Attention Heads, 125 millions of parameters, 2 epochs of additional training on Russian texts about medications;
- XLM-RoBERTa-large-sag—24 Transformer blocks, 1024 hidden neurons, 16 Attention Heads, 355 millions of parameters, 1 epoch of additional training on Russian texts about medications;

Text preprocessing includes splitting it into words or word parts—"tokens". For XLM-RoBERTa-sag, as well as for the original XLM-RoBERTa, such splitting is performed using the SentencePiece tokenizer [37].

Input Text Pre-Processing

To solve the classification task, transformer-based language models use a special token [CLS] added to the input sequence. During training, the loss functionn is aimed at class prediction based on the vector of the [CLS] token. That way, the model learns to create such a vector representation of the [CLS] token that accumulates information about the text as a whole and is informative in terms of the current task being solved.

In the approach proposed in this work, the classification is performed on the basis of the information about a pair of entities for which the existence of a relationship is determined, and the text that mentions this pair. Figure 2 shows the conceptual scheme of our approach to solving the relation extraction task using a language model.

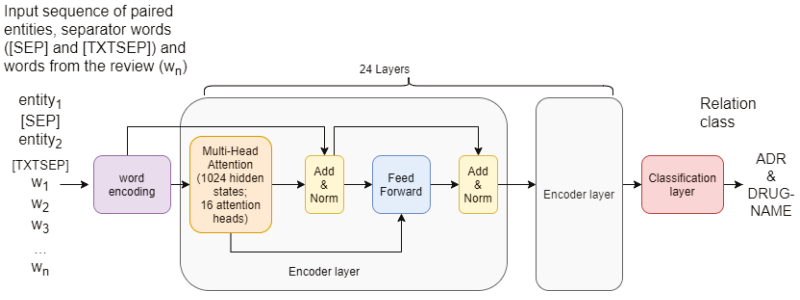


Figure 2. Conceptual scheme of our approach to relation extraction based on a language model.

For providing the language model with the information about which entities are of interest, several text representation variants are considered in our experiments:

1. The whole text—the tokenized input text that the language model receives at its input is the whole drug review text, in which target entities are highlighted using special start and end tokens, e. g. [T_ADR] and [\T_ADR] for an entity of type ADR:
   «[CLS]Antiviral syrup for children [T_DRUG]"Orvirem"[\T_DRUG] - We have an allergy to it! We have a severe allergy after the first day of taking it. Moreover, the boy (3.5 years old) had no allergies to any drugs before. In the morning he woke up covered in [T_ADR]red spots[\T_ADR]. I immediately gave him zyrtek... »
2. The text of target entities only—only the mentions of the target entities are used as the input text;
3. The text of the target entities and the text between the mentions of the target entities;
4. The text of the target entities concatenated with the whole text:
   [CLS]«text of first target entity»[SEP]«text of second target entity»[TXTSEP]«whole text of the drug review».

Here, the token [SEP] is placed between the two target entities, and the token [TXTSEP] separates the pair of entities from the whole text.

Potentially, this way of organizing the input data makes it possible to build a more informative vector representation due to the Attention mechanism inside the Transformer layers, and facilitates solving the problem in a classification formulation. The effectiveness of such a text representation was demonstrated previously [38].

As mentioned before, there are many degrees of freedom in such models that require consideration in order to achieve higher accuracy. Within the scope of the current research, the following options have been analyzed:

- Maximum input sequence length (in tokens);
- Learning rate;
- Batch size;
- Maximum learning epoch number;
- Learning rate decay technique [39];
- Early stopping technique [40].

3.2.2. Other Machine Learning Methods

Basic machine learning methods perform quite effectively in many applications [41–43]. These methods are highly efficient in terms of computational complexity, due to this fact, it is possible to search for the optimal set in an extensive space of hyperparameters and to test hypotheses relatively quickly.

The first goal of using basic machine learning methods was to obtain a reasonable baseline for the relation extraction task in the pharmacological domain in the Russian language, exceeding the random guess of the "Dummy" models' results, for the purpose of comparison with the deep learning models described in the previous section.

As a textual data representation for the baseline methods, concatenation of frequency features (term frequency-inverse document frequency, TF-IDF) of the character n-grams of the target entities was used. The size of the n-gram n and the frequency filter of the tf-idf method were considered as the hyperparameters to tune during the experiments.

The second goal of using basic machine learning methods was to check if the information about the entities' text is sufficient to achieve competitive accuracy for the task.

The following methods were considered in the experiments as basic machine learning methods:

- Logistic regression [44]—a basic linear model for text classification using a logistic function to estimate the probability of an example to belong to a certain class;
- Support vector machine [45]—a linear model based on building a hyperplane that maximizes the margin between two classes;
- Multinomial Naive Bayes model [46]—a popular solution for baselines in such text analysis tasks as spam filtering or text classification. It performs text classification based on words' or n-grams' co-occurrence probability;
- Gradient Boosting [47]—a strong decision tree-based ensemble model, which iteratively "boosts" the result of each tree by building a next tree that should classify examples that the previous tree fails to classify correctly.

In addition, for comparison, the RuBERT [48] language model was considered, which is a BERT [49] model with 12 layers, 768 hidden neurons each, 12 attention heads, 180 M parameters. RuBERT was trained on the Russian part of Wikipedia and news data. When solving the problem, the language model is used to form a vector representation of the text, which is fed into the linear layer. The output activities of the linear layer are used to determine if there is a relationship between the pair of entities fed to the input.

3.2.3. Dummy Models

"Dummy" models were considered to be the low-level baseline. Such models generate labels randomly or according to some simple principle. The following methods were checked as methods for "dummy" classification:

- most frequent class labeling—every pair of entities is assigned to the most frequent class in the dataset (in case of extraction of ADR–DrugName relations in the RDRS dataset, thus classifier considers every pair to have a relation);
- uniform random labeling—labels are predicted randomly according to a uniform probability distribution, without taking into account any characteristics of the input dataset;
- stratified random labeling—labels are predicted randomly but from the distribution corresponding to that of the input data: the probability of an input example to belong to a class is proportional to the portion of examples of such class in the dataset.

The accuracy of the "dummy" methods based on the random label generation was averaged over 100 launches in order to operate with more stable results and to prevent possible occurrence of random outliers.

4. Experiments

4.1. Accuracy Metric

The performance of a model on the relation extraction task is estimated by the f1-macro metric, in which the f1 score is calculated separately for each class:

$$\text{f1score} = \frac{2 \cdot P \cdot R}{P + R}, \tag{1}$$

$$P = \frac{TP}{TP + FP}, \tag{2}$$

$$R = \frac{TP}{TP + FN}. \tag{3}$$

Here P is precision, the proportion of correctly predicted objects of the class A under consideration to the number of objects that the model assigned to class A; R is recall, the proportion of correctly predicted items of class A to the real number of items in class A; TP is the number of *true positive* instances, the number of relations of class A correctly identified by the model; FP is the number of *false positive* examples, the number of relations assigned to class A while actually having a different class; FN is *false negative*, the number of relations that actually have class A while being incorrectly assigned to a different class by the model.

The overall performance of the model is estimated by averaging the f1-score over the two classes. This method of averaging allows for the uneven numbers of relations in the different classes.

4.2. Selection of the Model Features and Hyperparameters

In these experiments we use a subset of RDRS that contains texts with the ADR and Drugname entities only. The following experimental setup is used:

- Fixed stratified split into training (80%) and testing (20%) sets; In order to avoid overfitting, entity pairs from each review all go either to the training set or to the testing set, but no review is split between the sets;
- Hyperparameters of the language model's fine-tuning process are searched manually so that to maximize the accuracy (by the f1-macro metric) on the validation part of the training set, without taking into account the testing set;
- The language model involves early stopping and learning rate decay (Experiments show the positive effect of such techniques on the model accuracy);

The experiments on language models have been carried out using a computing cluster node with the following configuration: CPU Intel® Xeon™ E5-2650v2 (2.6 GHz) × 8, 128 Gb RAM, NVIDIA Tesla V100 (16 Gb).

4.3. An Estimation of Efficiency of Selected Methods

After finding the optimal model parameters, the efficiency of the methods has been assessed on a part of the RDRS containing review texts with multiple contexts. Accuracy is measured by the f1-score metric with cross-validation over 5 splits: the data is divided into 5 equal parts, and at each iteration of the cross-validation 80% of the texts are used for fine-tuning the model and 20% for testing.

For a more complete assessment, we compare the proposed method to other machine learning methods different in terms of complexity and type, as well as to a "Dummy" classifier based on the probability distribution of positive and negative examples of the pairs of entities in question (Stratified random labeling).

"Dummy" models and basic machine learning method experiments have been carried out on a local machine with the following configuration: CPU Intel® Core™ i5-7400 @ 3.00 GHz × 4, 16 Gb RAM. The experiments with language models were performed on the same equipment as the experiments in the previous section.

The programming language python 3.8 and software libraries numpy [50], sklearn [51], pytorch [52] and simpletransformers [53] were used for software implementation of the described method. As part of a series of experiments, the parameters of the python random number generator, as well as the random number generators of numpy, sklearn, and pytorch libraries were fixed to ensure repeatability of the experiments.

5. Results

5.1. Comparison of the Model Features and Hyperparameters

This section compares the results of experiments on the identification of the entity relations using XLM-RoBERTa-large-sag and XLM-RoBERTa-sag with different input representations described in Section 3.2.1.

Table 5. Accuracies (by the f1-macro metric) of XLM-RoBERTa-base-sag (denoted "LM-Base") and XLM-RoBERTa-large-sag ("LM-Large") language models with different methods of text representation.

Text Representation	LM-Base	LM-Large
Text of target entities only	0.75	0.76
Whole text with highlighting target entities	0.78	0.82
Text of target entities and text between them	0.81	0.80
Text of target entities and the whole text	0.91	0.95

The comparison shows that the language model should receive both the target entities separated from the text and the entire text in order to achieve high accuracy and to outperform basic machine learning methods. The f1-macro achieved for ADR–Drugname relations from the RDRS dataset is 95% (see Table 5). This estimation is 41% higher than random class prediction, and 20% higher than basic machine learning models, even if the hyperparameters of the latter are tuned on a test set.

The optimal hyperparameter values found for XLM-RoBERTa-base-sag are:

- maximum input length—512;
- early stopping—active;
- learning rate—0.00005;
- batch size—32;
- maximum epochs—10;
- learning rate decay—active;

The resulting hyperparameter values for XLM-RoBERTa-large-sag are:

- maximum input length— 512;
- early stopping—active;
- learning rate—0.00001;
- batch size—8 (there was not enough memory for bigger batch size with XLM-RoBERTa-large);
- maximum epochs—10;
- learning rate decay—active;

5.2. Estimation of the Relation Extraction Efficiency

As a result of the experiments conducted on the 908 reviews from the RDRS corpus that have multi-context annotation, accuracy has been estimated for the task of determining the relationships between pharmacologically-significant entities using the method developed on base of the XLM-RoBERTa language model. The accuracy of the proposed method in comparison with the baseline classifiers is given in Table 6. Accuracy is measured by the f1-score metric averaged over five cross-validation splits and is presented separately for the positive (relation present) class and the negative (no relation) class. The results for the baseline machine learning methods are obtained with input represented by target entity pairs encoded with tf-idf of n-grams of 3–8 characters.

As follows from this table, the proposed model determines the four relations under consideration with the following accuracy (according to the f1-score metric for the positive class): between adverse drug reactions and drugs (ADR–Drugname) 92.7%, between drugs and diseases (Drugname–Diseasename) 89.9%, between a drug and its source of information (Drugname–SourceInfoDrug) 92.9%, between diseases and symptoms (Diseasename–Indication) 87.1%. This is 43.5%, 40%, 41.5%, 38.2% higher than the accuracy of the dummy classifier and higher than the accuracy of RuBERT by 3.9%, 3.8%, 3.5%, 2.1% respectively. At the same time, for the class without the relation between entities (negative class), the accuracy is more volatile, taking the values of 91.1%, 76.2%, 82.7%, 31%. However, they exceed the Dummy Classifier accuracy by 59.3%, 42.9%, 49.8%, 9% and RuBERT by 14.9%, 10.0%, 20.1%, 3.3% respectively. On average, the developed model outperforms RuBERT by 3.6%, achieving the f1-score of 80.4%.

Table 6. Accuracy of predicting relations (pos) and absense of relations (neg) between entity pairs of different types in multicontext reviews from the RDRS dataset.

Methods	ADR–Drugname		Drugname–Diseasename		Drugname–Source Info Drug		Diseasename–Indication	
	pos	neg	pos	neg	pos	neg	pos	neg
Proposed model	92.7	91.1	89.9	76.2	92.9	82.7	87.1	31
	91.9	83.05	87.8	59				
RuBERT	88.8	76.2	86.1	66.2	89.4	72.6	85.7	27.7
	82.5	76.15	81	56.7				
LinearSVM	72.8	45.0	75.6	44.9	77.9	45.2	83.2	24.4
	58.9	60.25	61.55	53.8				
Multinomial Naive Bayes	66.3	33.8	68.8	26.1	73.4	14.3	80.2	5.4
	50.05	47.45	43.85	42.8				
Stratified Random Labeling	66.5	31.8	66.5	33.3	69.8	32.9	77.8	22.0
	49.15	49.9	51.35	49.9				

5.3. Applying the Proposed Approach to the DDI Dataset

In order to test the applicability of the proposed model to the texts in another language, we estimated our model on the well-known Drug-Drug Interaction (DDI) dataset [27], used on the SemEval-2013 challenge as a dataset for biomedical relation extraction.

The DDI dataset is a manually annotated corpus consisting of 792 texts selected from the DrugBank database and other 233 Medline abstracts. The dataset has been annotated with a total of 18,502 pharmacological substances and 5028 relations. The dataset includes named entities of the following types:

- Drug—used to annotate those human medicines known by a generic name;
- Brand—drugs described by a trade or brand name;
- Group—drug interaction descriptions often include groups of drugs, that were separated to "group" entity type;
- Drug_n— active substances that weren't approved for human use, such as toxins or pesticides.

The relations annotated in the dataset are four types of drug-drug interactions (DDIs):

- Mechanism— this type is used to annotate DDIs that are described by their pharmacokinetic mechanism;
- Effect—this type is used to annotate DDIs describing an effect or a pharmacodynamic mechanism;
- Advice—This type is used when a recommendation or advice regarding a drug interaction is given;
- Int—This type is used when a DDI appears in the text without providing any additional information.

When applying the proposed model to the DDI dataset, the model has been fine-tuned to DDI, but pre-training on Russian texts has remained the same. For the fine-tuning and testing on DDI, the data split is the same as in the BLURB project [18]—624/90/191 documents for train/validation/test sets respectively.

Experiments with the text representation "target entities and the whole text" (described in Section 3.2.1) yield the micro-averaged f1-score value of 71.2%. We have therefore modified the input text representation for higher accuracy on the DDI dataset. Inspired by the entity screening technique from the literature [18], we have employed both highlighting the target entity mentions with tags and concatenating target entity mentions with the texts of the whole reviews. For example, the text: "*Cytochrome P-450 inducers, such as phenytoin, carbamazepine and phenobarbital, induce clonazepam metabolism, causing an approximately*

30% decrease in plasma clonazepam levels." was represented as: *"phenytoin [SEP] carbamazepine [TXTSEP] Cytochrome P-450 inducers, such as [T]phenytoin[/T], [T]carbamazepine[/T] and phenobarbital, induce clonazepam metabolism, causing an approximately 30% decrease in plasma clonazepam levels."*

The resulting accuracy by f1-metric, micro-averaged over the four relation classes, is 81.46%, which is comparable to the accuracy other language model-based approcahes [18,20,54] achieve for determining relations between entities extracted from this dataset, the state of the art being 84.05% [20].

6. Discussion

Table 6 shows that there is a volatility between different relation types in terms of accuracy. Preliminary analysis of the relations of different types shows that the Diseasename–Indication relation has the following distinctive features: low number of the negative class samples (pairs of entities of the desired type that have no relation), high fraction of the unique pairs of entities (approx. 65%); high fraction (approx. 35%) of the unique relations that are represented with different classes in different texts (the same entity pair that has a relation in one text may have no relation in the other text). All these factors—unevenness of classes and the ambiguity of the relation existence between mentions of the same entities in different texts—make the classification task more difficult for the machine learning model. As a solution, we consider conducting further research of the data structure and the classification results, as well as extending our dataset by more relations that have lower representation in the corpus.

Overall, the developed approach shows the highest accuracy out of a group of methods considered: the language model RuBERT, trained on the Russian Wikipedia and news, classic machine learning algorithms (LinearSVM and Multinomial Naive Bayes) and the baseline"dummy" method of Stratified Random Labeling.

Though accuracy is the key performance value of the machine learning models, another important metric is their computational complexity. XLM-RoBERTa is the most resource-intensive model among those considered—it has approximately 550 million parameters, while RuBERT has approximately 110 million parameters. A limitation of XLM-RoBERTa-large is that it requires a GPU to work efficiently.

Another limitation of the transformer language models related to the computational complexity is a limit on the input sequence length—the input of the base BERT model cannot have a size larger than 512 tokens, while RoBERTa-large has this limitation set to 1024 tokens. In the case of longer texts, special approaches are needed in order to work efficiently and to use information about the whole text.

It is worth mentioning that this work considers the relation extraction task based on the ground truth named entity annotation, therefore, further research is required to determine the method's efficiency when the named entities are predicted by another model.

7. Conclusions

The research conducted shows the strong dependency of the accuracy of the entity relation identification on the structure of the input text representation when using pre-trained language models based on the Transformer topology. The highest accuracy is obtained with our proposed model XLM-RoBERTa-large-sag with texts represented in the following form: the text of the first entity of the potential relation, followed by the text of the second entity of the potential relation and the whole input text. The information contained inside the text between the target entities proved to be insufficient to achieve the same accuracy with the same model, presenting the entire review text to the language model is thus necessary.

The average f1-score obtained over 4 relation types is 80.4%. At the same time, the RuBERT model yields a 3.6% lower f1-score, Linear SVM—21.77% lower, baseline stratified random labeling method—30.4% lower.

On the DDI dataset the same model achieves 81.46%, which is comparable to the state-of-the-art 84.05% obtained on that dataset by other language models trained on pre-defined NER annotation.

Another important observation is the volatility of the accuracy across the relation types, which could be explained by the imbalance in the number of relations of different types, in the number of unique representatives of entity mentions, and the distribution of the relations of particular type between training and test subsets. The issue could be corrected with enlarging the corresponding parts of the context-labeled dataset and with balancing the numbers mentioned above.

Overall, the results obtained provide the state-of-the-art accuracy level for the task of pharmacological entity relation identification in Russian-language reviews and could be positioned as a basis for the future tasks of automated analysis of medical reviews.

Author Contributions: Conceptualization, A.S. (Alexander Sboev) and R.R.; methodology, A.S. (Alexander Sboev) and I.M.; software, A.S. (Anton Selivanov), G.R., I.M. and A.G.; validation, G.R. and A.S. (Anton Selivanov); investigation, A.S. (Alexander Sboev), A.S. (Anton Selivanov), G.R. and I.M.; resources, A.S. (Alexander Sboev) and R.R.; data curation, A.S. (Alexander Sboev), S.S. and A.G.; writing—original draft preparation, R.R., A.S. (Anton Selivanov) and A.S. (Alexander Sboev); writing—review and editing, A.S. (Alexander Sboev), R.R. and A.S. (Anton Selivanov); visualization, A.G. and G.R.; supervision, A.S. (Alexander Sboev); project administration, R.R.; funding acquisition, A.S. (Alexander Sboev). All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the Russian Science Foundation grant No. 20-11-20246.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be obtained through sending a request from the website of our project: https://sagteam.ru/en/med-corpus/ (accessed on 30 October 2021). Trained models are presented on the page of our team on the huggingface repository: https://huggingface.co/sagteam (accessed on 30 October 2021). The code is available at https://github.com/sag111/Relation_Extraction (accessed on 30 October 2021).

Acknowledgments: This work has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC "Kurchatov Institute", http://ckp.nrcki.ru/ (accessed on 30 October 2021)

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Segura-Bedmar, I.; Martínez, P. Pharmacovigilance through the development of text mining and natural language processing techniques. *J. Biomed. Inform.* **2015**, *58*, 288–291. [CrossRef] [PubMed]
2. Sboev, A.; Sboeva, S.; Gryaznov, A.; Evteeva, A.; Rybka, R.; Silin, M. A neural network algorithm for extracting pharmacological information from russian-language internet reviews on drugs. *J. Phys. Conf. Ser.* **2020**, *1686*, 012037. [CrossRef]
3. Sboev, A.; Sboeva, S.; Moloshnikov, I.; Gryaznov, A.; Rybka, R.; Naumov, A.; Selivanov, A.; Rylkov, G.; Ilyin, V. An analysis of full-size Russian complexly NER labelled corpus of Internet user reviews on the drugs based on deep learning and language neural nets. *arXiv* **2021**, arXiv:cs.CL/2105.00059.
4. Oliveira, A.; Braga, H. Artificial Intelligence: Learning and Limitations. *Wseas Trans. Adv. Eng. Educ.* **2020**, *17*, 80–86. [CrossRef]
5. Al-Haija, Q.A.; Jebril, N. A Systemic Study of Pattern Recognition System Using Feedback Neural Networks. *Wseas Trans. Comput.* **2020**, *19*, 115–121. [CrossRef]
6. Ganesh, P.; Rawal, B.; Peter, A.; Giri, A. POS-Tagging based Neural Machine Translation System for European Languages using Transformers". *Wseas Trans. Inf. Sci. Appl.* **2021**, *18*, 26–33. [CrossRef]
7. Xu, H.; Van Durme, B.; Murray, K. BERT, mBERT, or BiBERT? A Study on Contextualized Embeddings for Neural Machine Translation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 6663–6675.
8. Ge, Z.; Sun, Y.; Smith, M. Authorship attribution using a neural network language model. In Proceedings of the AAAI Conference on Artificial Intelligence, Burlingame, CA, USA, 8–12 October 2016; Volume 30.

9. Peters, M.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K. Deep contextualized word representations. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; Volume 1.
10. Luong, M.T.; Pham, H.; Manning, C.D. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:1508.04025.
11. Portelli, B.; Passabi, D.; Serra, G.; Santus, E.; Chersoni, E. Improving Adverse Drug Event Extraction with SpanBERT on Different Text Typologies. In Proceedings of the 5th International Workshop on Health Intelligence (W3PHIAI-21), Palo Alto, CA, USA, 8–9 February 2021.
12. Yan, H.; Gui, T.; Dai, J.; Guo, Q.; Zhang, Z.; Qiu, X. A Unified Generative Framework for Various NER Subtasks. *arXiv* **2021**, arXiv:2106.01223.
13. Ge, S.; Wu, F.; Wu, C.; Qi, T.; Huang, Y.; Xie, X. FedNER: Privacy-Preserving Medical Named Entity Recognition with Federated Learning. Available online: https://arxiv.org/abs/2003.09288 (accessed on 30 October 2021).
14. Wu, S.; He, Y. Enriching pre-trained language model with entity information for relation classification. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2361–2364.
15. Giorgi, J.; Wang, X.; Sahar, N.; Shin, W.Y.; Bader, G.D.; Wang, B. End-to-end named entity recognition and relation extraction using pre-trained language models. *arXiv* **2019**, arXiv:1912.13415.
16. Eberts, M.; Ulges, A. Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training. In *ECAI 2020*; IOS Press: Amsterdam, Netherlands,2020; pp. 2006–2013.
17. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2019**, *36*, 1234–1240. [CrossRef]
18. Gu, Y.; Tinn, R.; Cheng, H.; Lucas, M.; Usuyama, N.; Liu, X.; Naumann, T.; Gao, J.; Poon, H. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *arXiv* **2020**, arXiv:2007.15779.
19. Gordeev, D.; Davletov, A.; Rey, A.; Akzhigitova, G.; Geymbukh, G. Relation extraction dataset for the russian language. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog" [Komp'iuternaia Lingvistika i Intellektual'nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii "Dialog"]*; Russian State University For The Humanities: Moscow, Russia, 2020.
20. Naseem, U.; Dunn, A.G.; Khushi, M.; Kim, J. Benchmarking for biomedical natural language processing tasks with a domain specific albert. *arXiv* **2021**, arXiv:2107.04374.
21. Ju, M.; Nguyen, N.T.; Miwa, M.; Ananiadou, S. An ensemble of neural models for nested adverse drug events and medication extraction with subwords. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 22–30. [CrossRef]
22. Joshi, M.; Chen, D.; Liu, Y.; Weld, D.S.; Zettlemoyer, L.; Levy, O. Spanbert: Improving pre-training by representing and predicting spans. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 64–77. [CrossRef]
23. Wang, J.; Lu, W. Two Are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 1706–1721.
24. Patrick, J.; Li, M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. *J. Am. Med. Inform. Assoc.* **2010**, *17*, 524–527. [CrossRef]
25. Anick, P.; Hong, P.; Xue, N.; Anick, D. I2B2 2010 challenge: Machine learning for information extraction from patient records. In *Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data*; Boston, MA, USA, 12 November 2010.
26. Henry, S.; Buchan, K.; Filannino, M.; Stubbs, A.; Uzuner, O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. *J. Am. Med. Inform. Assoc.* **2019**, *27*, 3–12. [CrossRef]
27. Herrero-Zazo, M.; Segura-Bedmar, I.; Martínez, P.; Declerck, T. The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *J. Biomed. Inform.* **2013**, *46*, 914–920. [CrossRef]
28. Asada, M.; Miwa, M.; Sasaki, Y. Using Drug Descriptions and Molecular Structures for Drug-Drug Interaction Extraction from Literature. *Bioinformatics* **2020**, *37*, 1739–1746. [CrossRef]
29. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: Pretrained Language Model for Scientific Text. *arXiv* **2019**, arXiv:1903.10676.
30. Gurulingappa, H.; Rajput, A.M.; Roberts, A.; Fluck, J.; Hofmann-Apitius, M.; Toldo, L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *J. Biomed. Inform.* **2012**, *45*, 885–892. [CrossRef]
31. Bruches, E.; Pauls, A.; Batura, T.; Isachenko, V. Entity Recognition and Relation Extraction from Scientific and Technical Texts in Russian. In Proceedings of the 2020 Science and Artificial Intelligence Conference (SAI Ence), Novosibirsk, Russia, 14–15 November 2020; pp. 41–45.
32. Ivanin, V.; Artemova, E.; Batura, T.; Ivanov, V.; Sarkisyan, V.; Tutubalina, E.; Smurov, I. Rurebus-2020 shared task: Russian relation extraction for business. In *Computational Linguistics and Intellectual Technologies*; Russian State University for the Humanities: Moscow, Russia, 2020; pp. 416–431.
33. Bondarenko, I.; Berezin, S.; Pauls, A.; Batura, T.; Rubtsova, Y.; Tuchinov, B. Using Few-Shot Learning Techniques for Named Entity Recognition and Relation Extraction. In Proceedings of the 2020 Science and Artificial Intelligence Conference (SAI Ence), Novosibirsk, Russia, 14–15 November 2020; pp. 58–65.

34. Loukachevitch, N.; Artemova, E.; Batura, T.; Braslavski, P.; Denisov, I.; Ivanov, V.; Manandhar, S.; Pugachev, A.; Tutubalina, E. NEREL: A Russian Dataset with Nested Named Entities and Relations. *arXiv* **2021**, arXiv:2108.13112.
35. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised cross-lingual representation learning at scale. *arXiv* **2019**, arXiv:1911.02116.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.U.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 5998–6008.
37. Kudo, T.; Richardson, J. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv* **2018**, arXiv:1808.06226.
38. Sboev, A.; Selivanov, A.; Rybka, R.; Moloshnikov, I.; Rylkov, G. Evaluation of Machine Learning Methods for Relation Extraction Between Drug Adverse Effects and Medications in Russian Texts of Internet User Reviews. Available online: https://pos.sissa.it/410/006/pdf (accessed on: 30 October 2021).
39. Smith, L.N. Cyclical learning rates for training neural networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 464–472.
40. Caruana, R.; Lawrence, S.; Giles, L. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Adv. Neural Inf. Process. Syst.* **2000**, *13*, 402–408.
41. Sahoo, K.S.; Tripathy, B.K.; Naik, K.; Ramasubbareddy, S.; Balusamy, B.; Khari, M.; Burgos, D. An evolutionary SVM model for DDOS attack detection in software defined networks. *IEEE Access* **2020**, *8*, 132502–132513. [CrossRef]
42. Chun, P.J.; Izumi, S.; Yamane, T. Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine. *Comput.-Aided Civil Infrastruct. Eng.* **2021**, *36*, 61–72. [CrossRef]
43. Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Inf. Process. Manag.* **2020**, *57*, 102221. [CrossRef]
44. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398.
45. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]
46. Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 workshop on empirical methods in artificial intelligence, Seattle, WA, USA, 4 August 2001; Volume 3, pp. 41–46.
47. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. Boosting algorithms as gradient descent in function space. In Proceedings of the NIPS, Denver, CO, USA, 29 November–4 December 1999; Volume 12, pp. 512–518.
48. Kuratov, Y.; Arkhipov, M. Adaptation of deep bidirectional multilingual transformers for Russian language. In *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*; Russian State University For The Humanities: Moscow, Russia, 2019; pp. 333–339.
49. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
50. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]
51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
52. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32, pp. 8026–8037.
53. Rajapakse, T.C. Simple Transformers. 2019. Available online: https://github.com/ThilinaRajapakse/simpletransformers (accessed on 30 October 2021).
54. raj Kanakarajan, K.; Kundumani, B.; Sankarasubbu, M. BioELECTRA: Pretrained Biomedical text Encoder using Discriminators. In Proceedings of the 20th Workshop on Biomedical Language Processing, Online, 11 June 2021; pp. 143–154.

big data and cognitive computing

MDPI

Article

Gender Stereotypes in Hollywood Movies and Their Evolution over Time: Insights from Network Analysis

Arjun M. Kumar, Jasmine Y. Q. Goh, Tiffany H. H. Tan and Cynthia S. Q. Siew *

Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Singapore 117570, Singapore; arjun.k@u.nus.edu (A.M.K.); e0324141@u.nus.edu (J.Y.Q.G.); tiffanytan@u.nus.edu (T.H.H.T.)
* Correspondence: cynthia@nus.edu.sg

Abstract: The present analysis of more than 180,000 sentences from movie plots across the period from 1940 to 2019 emphasizes how gender stereotypes are expressed through the cultural products of society. By applying a network analysis to the word co-occurrence networks of movie plots and using a novel method of identifying story tropes, we demonstrate that gender stereotypes exist in Hollywood movies. An analysis of specific paths in the network and the words reflecting various domains show the dynamic changes in some of these stereotypical associations. Our results suggest that gender stereotypes are complex and dynamic in nature. Specifically, whereas male characters appear to be associated with a diversity of themes in movies, female characters seem predominantly associated with the theme of romance. Although associations of female characters to physical beauty and marriage are declining over time, associations of female characters to sexual relationships and weddings are increasing. Our results demonstrate how the application of cognitive network science methods can enable a more nuanced investigation of gender stereotypes in textual data.

Keywords: gender stereotypes; story tropes; movie plots; network analysis; word co-occurrence network

Citation: Kumar, A.M.; Goh, J.Y.Q.; Tan, T.H.H.; Siew, C.S.Q. Gender Stereotypes in Hollywood Movies and Their Evolution over Time: Insights from Network Analysis. *Big Data Cogn. Comput.* **2022**, *6*, 50. https://doi.org/10.3390/bdcc6020050

Academic Editor: Carson K. Leung

Received: 17 January 2022
Accepted: 13 April 2022
Published: 6 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Stereotypes are defined as "cognitive structures that provide knowledge, beliefs and expectations about individuals based on their social group membership" [1]. Being cognitive in nature, stereotypes can affect specific social perceptions of others, such as their personality, behaviour, attitudes, and appearance. For example, stereotypes can lead us into subscribing to the belief that women are communal by nature, which they display through being warm and sensitive to others, and the belief that men are agentic by nature, which they demonstrate by being independent and assertive in the presence of others [2]. It has been shown that the stereotypical categorisation of people into different groups is fluid and dependent on the context of comparisons [3]. However, gender classification seems to evade such fluid categorisation, since it is a primary and salient feature of the perception of other people [4]. Such immediately recognised and chronically salient categorisations contribute to the persistence of gender stereotypes. While stereotypes' cognitive nature encompasses the beliefs and descriptions that people hold about the members of different groups (e.g., gender), evaluations that follow these implicit or explicit attitudes could involve negative or positive reactions to members of a specific group. In the case of gender stereotypes, they are organised around the importance of the agency of men and the communality of women. Therefore, task performance is emphasised for men, whereas social relationships are emphasised for women [5]. Women who violate the prescriptive gender stereotype of being warm and kind in social relationships could potentially face backlash for acting against these prescribed gender norms. Negative evaluations of gender-stereotype violations could result in discriminatory behaviour.

However, much has changed with respect to the roles and expectations of genders in recent years. In the United States, gender employment and pay gaps have reduced

substantially since the enactment of the Equal Pay Act in 1963 [6]. More women are entering science, technology, engineering, and mathematics (STEM) fields, taking on leadership roles, and making key scientific discoveries than before [7]. There have also been four major waves of feminism. with each one marking a specific cultural era and women's involvement with the media [8]. In addition, fewer women are marrying and those do marry do so at a much older age compared to the past [9].

Social role theory advocates that stereotypes arise from observations of real-life behaviour; in other words, they represent our lived realities. Thus, the changing roles of men and women in society should also influence stereotypes in society [10]. In fact, this is what has been observed across numerous studies and meta-analyses based on questionnaires on implicit and explicit attitudes [11–13]. More recently, Charlesworth and Banaji [14] found decreasing gender stereotypes (male—science/female—arts, male—career/female—family) from measures of implicit and explicit attitudes across a 10-year period from 2007 to 2018. However, these are indicators of subjective individual attitudes in a controlled setting, and may be limited in their representation of implicit attitudes of a society [15]. Cultural products such as movies may complement individual perceptions to offer a richer account of how gender stereotypes prevail in society, especially since they do not exist independently from one another. An individual's perception and acceptance of a stereotype into their schema is influenced by their experiences—including those shaped by the media [16]—while the media products a culture creates are ultimately products of the same perceptions that individuals hold [17]. Therefore, both individual attitudes and cultural products make for apt sources through which to study the evolution of gender stereotypes in society, and ought to supplement each other.

Hence, Durkheim [18] and his proponents argued that the primary materials through which to investigate social stereotypes are the products of a society. This is in contrast to the psychological studies mentioned previously, which typically ask participants to complete questionnaires or take part in laboratory experiments in order to measure the types and strength of gender stereotypes. Durkheim's idea of studying the cultural products of society on a large scale was not feasible during his time, but it is possible today given the drastic increase in computational power along with the accessibility of big data. For example, Stella's paper [19] used textual forma mentis networks (TFMN) to reconstruct large-scale online discourses regarding the gender disparity in STEM fields on the social media platform Twitter, and the researchers found largely positive, gender-stereotype-free perceptions of the gender gap. Other studies used word-embedding models to study gender stereotypes and their dynamic changes across time [7,20]. Word embeddings are machine-learning methods in which the words in a language are represented using high-dimensional vectors. Geometric relationships between the vectors denote semantic relationships between the words. For example, words that are geometrically closer to each other in the vector space are semantically closer to each other [21]. Bhatia and Bhatia [20] used word embeddings to examine a century's worth of shifts and developments in the gender biases expressed in large-scale historical natural-language data, and found that these stereotypes decreased in strength over time. It is important to note that the changes they observed seem to have been driven mainly by changes in stereotypically feminine traits (as opposed to stereotypically masculine ones) and personality-related traits (as opposed to physical traits). Charlesworth et al. [7] analysed the embeddings in more than 65 million words in English-language texts to investigate the presence of gender stereotypes and mainly found that gender stereotypes were pervasive and consistent across different age groups, sources, and time periods. However, they further noted that gender stereotypes that worked to the disadvantage of the different groups were not as prominently displayed in later time periods compared to earlier ones, indicating a trend towards more equitable representations. This meant that more recent corpora included fewer stereotypical gender associations, such as those of 'male–work' and 'female–home', depicting a reduction in the portrayal of women in caregiver roles and signifying a change away from the distinct segregation of traditional roles played by men and women in society.

Despite the usefulness of word embeddings in understanding semantic representations, network science offers unique perspectives in understanding cognitive structures such as semantic memory and the mental lexicon, or in this case, the representation of gender in the media products of a society. Network science can model dynamic changes in systems, as well as enabling the explicit investigation of the underlying structure of a model, something that is more challenging to accomplish with models derived from machine-learning methods [22]. Therefore, network science provides a compelling avenue through which one could represent and model changes in the implicit attitudes of a society. In one article employing a network science approach to studying stereotypes in movies, Xu et al. [23] studied gender stereotypes in 6087 movie synopses from the Internet Movie Database (IMDb) through a network analysis of word co-occurrences. They found that male lead characters were more commonly associated with verbs compared to female lead characters. By using community detection methods, they also showed that the key difference between the roles of leading male and female characters in movies was that women's role in romance is emphasised, whereas men's role in crime is emphasised. However, they did not study how these stereotypes changed or evolved across time. They also restricted their analysis to lead characters in movies.

Another avenue that could be explored is the analysis of narratives through story tropes in movies, which can be facilitated by network analysis. Barthes and Duisit [24] advocated that 'there has never been anywhere, any people without narrative'. A trope is a form of narrative that is viewed as a 'commonly recurring literary and rhetorical devices' [25]. In the case of movies, tropes refer to plot types of plot that audiences would expect a movie to tell, according to its theme or genre. An example would be the trope of 'enemies-to-lovers' in a romance-themed movie, describing a romantic arc whereby a couple who begin their relationship as enemies progressively fall in love. Tropes can also take the form of plot or narrative setups (e.g., a 'groundhog day' loop) or character types (e.g., a supervillain who avenges his parents' tragic death) ("Trope", 2020). In short, they are recognisable storytelling conventions, and can be understood as short, abbreviated, and decontextualised narratives within movies. Identifying tropes and examining how they change over time could thus be an important part of understanding society [26].

Our study builds on the aforementioned work by using a network-science perspective to understand stereotypes and their dynamic evolution in Hollywood movies (i.e., movies produced in North America). Given that our study is confined to Hollywood movies, the findings of our study are limited to stereotypes within a North American context. However, an overarching aim of this paper is to show researchers that the representation of genders and stereotypes could be modelled as a cognitive network. This allows us to add an additional tool to the pre-existing toolkit of methods within natural language processing. A network perspective allows not only the study of specific associations at the level of words, but also the overall structure of a network model. At the level of words, we investigated specific stereotypical associations in Hollywood movies by analysing specific edges within our network. At a broader level of the network model, we analysed stereotypical themes in Hollywood movies using the network's community structures. Additionally, by investigating how gender stereotypes change over time, we were able to examine whether changes in society may be reflected in the changes in stereotypes, which was not explored in the analysis by Xu et al. [23]. In contrast to their study, we also extended our analysis to include both lead and supporting characters in movies, thus providing a more complete picture of gender representations.

In the present study, we first examined the community structure of male and female characters' co-occurrence networks, which were constructed based on a frequency measure similar to that used in the study by Xu et al. [23]. Next, we identified story tropes associated with male and female characters using a novel method of path analysis from co-occurrence networks built using loglikelihood measures. Subsequently, we examined how the significance measures of stereotypical story tropes changed with time. Lastly, we identified the

most significant nouns, verbs, and adjectives in the networks and investigated how their edge weights in the network changed across time.

2. Materials and Methods

2.1. Description of Data Source

In total, 200 movies were randomly chosen for each year between 1940 and 2019 from Wikipedia's list of Hollywood films by year (e.g., https://en.wikipedia.org/wiki/List_of_American_films_of_2000, accessed on 9 November 2021). The 'Plot' sections of the Wikipedia entries of the chosen films were scraped using the 'rvest' R library [27]. Unlike Xu et al. [23], who used movie synopses from IMDb to construct their network, we used data from Wikipedia for two main reasons. Firstly, Wikipedia is a community-maintained database free for anyone to edit. Through the process of editing and re-editing by various contributors, the content on its site ultimately reaches an implicit consensus, i.e., when the current state of information is no longer disputed or corrected [28]. Therefore, the data on Wikipedia are typically representative of various voices, making the study of the implicit attitudes of a society viable. Secondly, plots contain more details about movies as compared to synopses, which are more akin to brief summaries of movies.

As mentioned above, for each year between 1940 and 2019, 200 movies were chosen randomly. However, each decade had a varying amount of data because of differences in the number of movies with a plot section, as well as differences in the average number of sentences in each movie plot. To counter the imbalance in the data for each decade and to maximise the available information at the same time, the decade with the least number of sentences was first identified. Next, we used the number of sentences in this decade to randomly sample across the data from each decade. The minimum number of sentences was found in the decade from 1940 to 1949, with 22,638 sentences. Hence, 22,638 sentences were randomly sampled from each decade. This resulted in a total of 181,104 sentences for the analysis, which were randomly selected from all the sentences in 16,000 movies.

The words in this dataset were then tokenised and parsed through natural language processing using the 'spacyr' R library [29] to tag part of speech information (e.g., kill/verb, John/person, etc.). 'Spacyr' is an R wrapper around the Python package called 'spacy', which is an open source library for advanced natural language processing [30] Subsequently, only character names and the grammatical categories of verbs, nouns, and adjectives were included, prepositions, function words, determiners, and symbols were removed. The 'genderizeR' R package [31] was then used to identify the gender of characters in the dataset. The 'genderizeR' package uses first names to predict gender using a database based on online social media profiles and their associated genders. An advantage of this package is that it is updated every day using information from new social media profiles. This method offers a high degree of accuracy because it is based on real-world data and is not a prediction-based model trained on a data set. For example, as of April 2015, the database contained 212,252 unique names gathered from 2 million social media profiles [32] Finally, character names were replaced with the general terms 'male/character' and 'female/character' based on their gender, as tagged by the 'genderizeR' R package. An example of the data processing is as follows: "Mary went to meet her father" becomes 'female/character' 'went/verb' 'meet/verb' 'father/noun'.

2.2. Network Construction

The graphs in our analysis were constructed using the 'igraph' R package [33] and plotted using the 'visNetwork' R package [34]. Using the processed dataset, we constructed two types of network for our analysis: (i) A raw-frequency-based network to understand the common themes around male and female characters using community detection, and (ii) a log-likelihood-test-based network to more specifically understand the lives of male and female characters in movies using edge weights of words linked to them. More details about the network construction for each of the networks are provided in specific sections below.

2.2.1. Common Themes for Male and Female Characters in Movies

To investigate the common themes that emerged in the words surrounding male and female characters, we identified communities in the frequency-based co-occurrence networks for male and female characters separately. Our approach was similar to the one used by Xu et al. [23], who identified community structures in the co-occurrence networks for male and female protagonists in Hollywood movies. However, our analysis was different in that we analysed communities from a network that considered the contexts of all male and female characters in movie synopses and not just the protagonists. First, we separated the dataset for male and female characters by selecting five words before and after the character names for males and females. We employed the same size of word window as Xu et al. [23] so we could standardise and compare the results of our analysis with theirs. The edges in the network were then weighted based on how frequently each pair of words appeared together within the same sentence across the entire corpus. Community detection was performed on this network using the Louvain algorithm [35] for both male and female character networks. The Louvain algorithm is a greedy optimisation method that attempts to optimise modularity when extracting community structures from large networks.

2.2.2. The Lives of Male and Female Characters in Movies

To investigate gender stereotypes in story tropes, we created a network in which the weight of edges was the log-likelihood of co-occurrence rather than frequency of co-occurrence, as used in the previous network. Rankings of most significant word pairs provide an indicator of their co-occurrence in word usage [36]. The log-likelihood test is essentially the use of a generalised likelihood ratio λ to compare two parameterised distributions (binomial in this case). One of the distributions is based on from the independence assumption, and the other from the observed frequencies. Taking $-2\log\lambda$, the significance value is obtained, which is χ^2-distributed. In their evaluation of the different measures of word similarity, such as baseline frequency, dice, and mutual information, Bordag [36] found that the significance of log-likelihood was the best. Hence, this measure was used in our analysis.

To construct the network, the male and female character vertices (henceforth referred to as gender vertices) were connected to the 20 most significant word associations in the movie plots based on the significance of the log-likelihood ratio of their co-occurrence within the same sentence. These word associations are referred to as primary associations of the gender vertices.

The significance of the log-likelihood ratio of co-occurrence was calculated as described in the Appendix (see Appendix A). Each of these primary associations with character names was then connected to its 20 most significant associations. These are henceforth referred to as secondary associations of the gender vertices. The steps of the network construction are shown in Figure 1 by taking 10 associations as an example, illustrating networks of both primary and secondary co-occurrences for both male and female characters. Vertices with a degree of less than two (i.e., fewer than two edges) were removed for ease of visualisation.

(a) **Gender-specific story tropes and their evolution.** The main goal of this analysis was to find the most significant story tropes associated with male and female characters across the entire time period of analysis (i.e., from 1940 to 2019). To this end, we used a novel method involving path analysis to identify the most significant story tropes in movies. First, we identified the most significant association of each of the primary vertices. Next, we computed the (weighted) path length from the gender vertices to these secondary vertices, through a path described by: gender vertex–primary vertex–secondary vertex. This resulted in 3-tuples of vertices and their path weights, which showed how significant they were. For example, along the path described by 'female/characters–love/noun–fell/verb', the cumulative weight would be the sum of the log-likelihood ratio significances (i.e., the weights) along the path. In other words, it would be the sum of LL (female/characters–love/noun) + LL (love/noun–fell/verb). Rather than simply inferring story tropes from primary vertices alone, the

3-tuples with the addition of secondary vertices provided us with additional context to infer story tropes of each gender. Twenty significant character-specific paths were obtained for each character, one for each of their primary associations. Figure 2 depicts the ten most significant paths for each gender within the co-occurrence network.

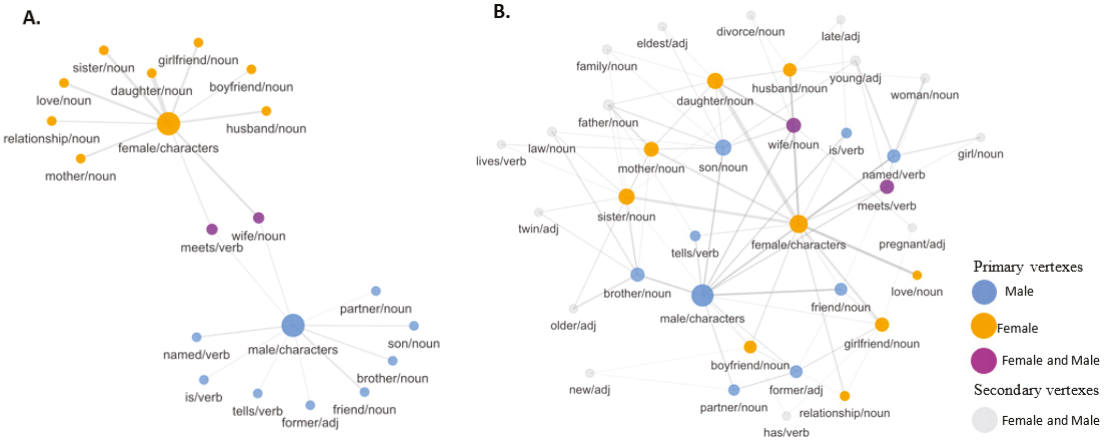


Figure 1. Details of the network construction. (**A**) Network with only primary nodes. Blue nodes represent unique associations with male characters, orange nodes represent unique associations with female characters, and purple nodes represent the common associations of both male and female characters. (**B**) Network with primary and secondary nodes. Secondary nodes are depicted in grey.

After identifying the top 20 tropes for each gender, we identified specific romance- and crime/violence-related tropes from this set, and analysed how their path weights changed across time for each decade from the 1940s to the 2010s, using linear regression. Romance and crime/violence were selected as gender-stereotypical domains worthy of further examination based on Xu et al.'s [23] findings. Xu et al.'s community analysis revealed eight gender-stereotypical themes or domains. Specifically, male characters were associated with crime, career, family, and action, and female characters with romance, career, family, and action. It should be noted that although Xu et al. [23] identified crime as a theme, we subsumed it under the more general theme of crime/violence. This is because an act linked to 'kill' does not necessarily refer to crime-related activities in all instances. For example, a male character could be involved in the killing of a monster rather than the killing of a human being, which does not definitively indicate that the male character has indeed participated in a crime. In Xu et al.'s study, most of the domains overlapped for both genders (i.e., both male and female characters were significantly associated with the career, family, and action stereotypes)—the only differences were that male characters were significantly associated with crime and female characters were not, whereas female characters were significantly associated with romance and male characters were not. As such, the 20 tropes were manually coded to identify specific tropes that fitted into these domains of crime/violence and romance for both genders for subsequent analysis. Importantly, the decision to include an analysis of the violence stereotype in female characters and romance in male characters despite these associations not being present in the communities allowed us to compare the presence and trends of these tropes between the genders across time.

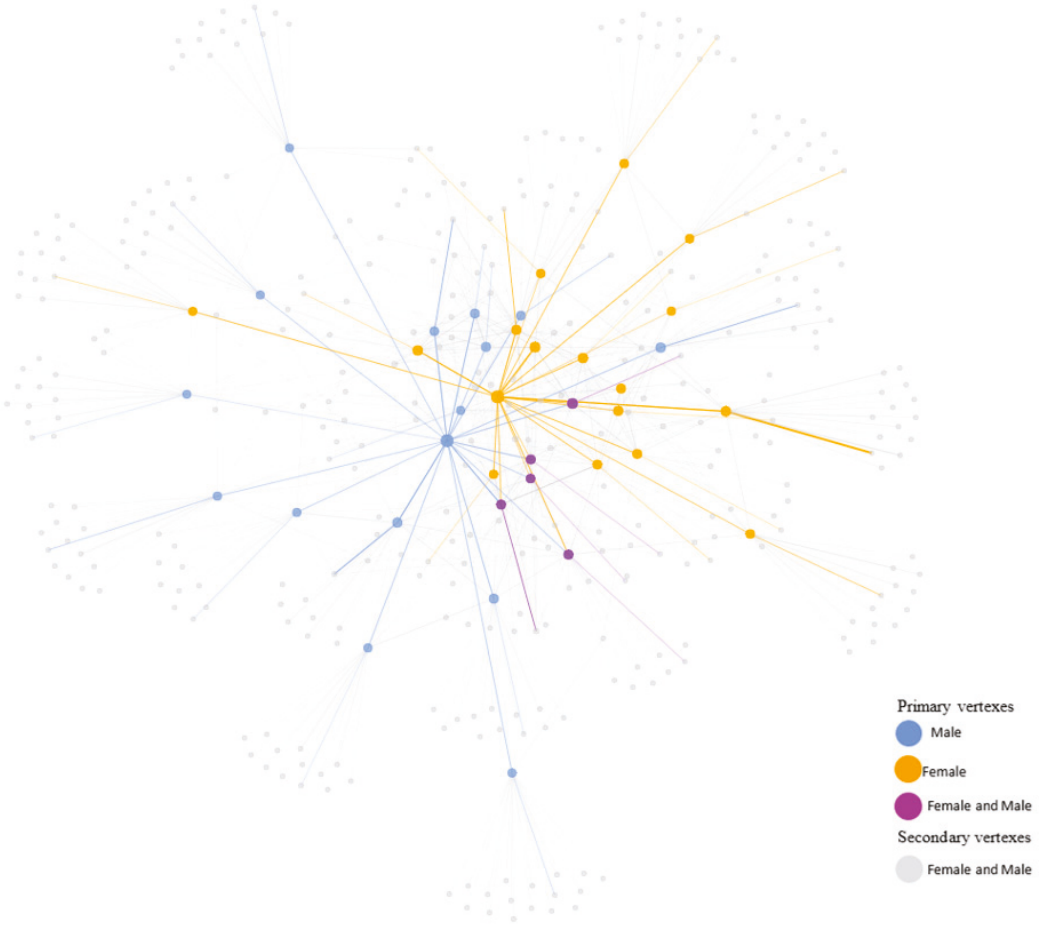


Figure 2. A representation of the ten most significant paths in the network for each gender among the whole network. Paths were selected after filtering out paths that crossed through the other gender vertex.

For instance, a 3-tuple that represents a romance trope from the network could be 'male–girlfriend–proposed' or 'female–fall–love', whereas a crime trope could be appear as 'male–kills–gun' or 'female–attacks–robbery'. After the relevant romance- and crime-related tropes were identified, each of their path weights were analysed for changes over time. This analysis enabled us to determine, for example, whether a trope such as 'male–kills–gun' became more or less prevalent between the 1940s and the 2010s.

(b) **Roles, actions, and descriptions of male and female characters and their evolution.** In this part, three network visualisations were created by subsetting the network formed previously based on three word classes: nouns, verbs, and adjectives. The vertices connected to edges of the gender vertices with the highest weight were analysed to find the most common associations for male and female characters. By analysing separate networks of nouns, verbs, and adjectives, we aimed to identify the top 20 most significant roles, actions, and descriptions, respectively, that were associated with male and female characters.

Furthermore, similar to our trope analysis, we sought to understand how stereotypical associations with individual words changed across time. Based on the same pre-identified domains of romance and crime/violence, we coded the individual nouns, verbs, and adjectives from the top 20 words and classified each as a crime/violence-related word or a romance-related word. For example, words such as 'love' or 'dating' were classified in the romance domain, whereas words such as 'felony' or 'prison' were classified in the crime/violence domain. After this categorisation, the edge weights of these associations were analysed for changes across decades using linear regression, which enabled us to determine whether a given word became increasingly or decreasingly associated with each gender over time (or whether there was no trend). To illustrate this, the co-occurrence of 'love' and female characters from 1940 to 2019 was analysed to determine whether female characters were increasingly associated with the word 'love' across that period, or less so. The same would apply to 'felony' or 'prison' in association with male characters. Ambiguous words that could not be clearly determined as being associated with either category were left out of this analysis.

3. Results

3.1. Common Themes for Male and Female Characters in Hollywood Movies

Five communities were identified for the male character network and the female character network, each using the Louvain algorithm [35]. The modularity of the female community structure was 0.07, and modularity of the male community structure it was 0.09. The low modularity values indicate that the community structure of these networks was not particularly robust. Using the top twenty vertices with the highest degree within each community (or all the vertices if the community had less than twenty vertices in total), we labelled each community with a specific theme; these are listed in Appendix B. We based our classification on the key words in the top twenty vertices and on the results presented by Xu et al. [23]. For the male network, the communities we identified were: crime, action, family, war, and plot narration. These communities are illustrated in Figure 3. The plot narration community is not an aspect of movies themselves, but instead reflects how movie plots are described on Wikipedia. For the female network, the communities identified were: crime, shopping, family, suicide, and plot narration. The number of vertices in each community is given in Table 1. The themes that are common to the lives of male and female characters in Hollywood movies are 'crime' and 'family'. Male characters differ in that they have the themes of 'war' and 'action'. Female characters differ in that they have the themes of 'shopping' and 'suicide'.

Table 1. Communities identified in the network and their total number of vertices.

a. Male Character		b. Female Character	
Community	**Number of Vertices**	**Community**	**Number of Vertices**
family	1465	crime	2876
action	1408	family	2547
war	1405	plot narration	1941
plot narration	1405	shopping	17
crime	524	suicide	12

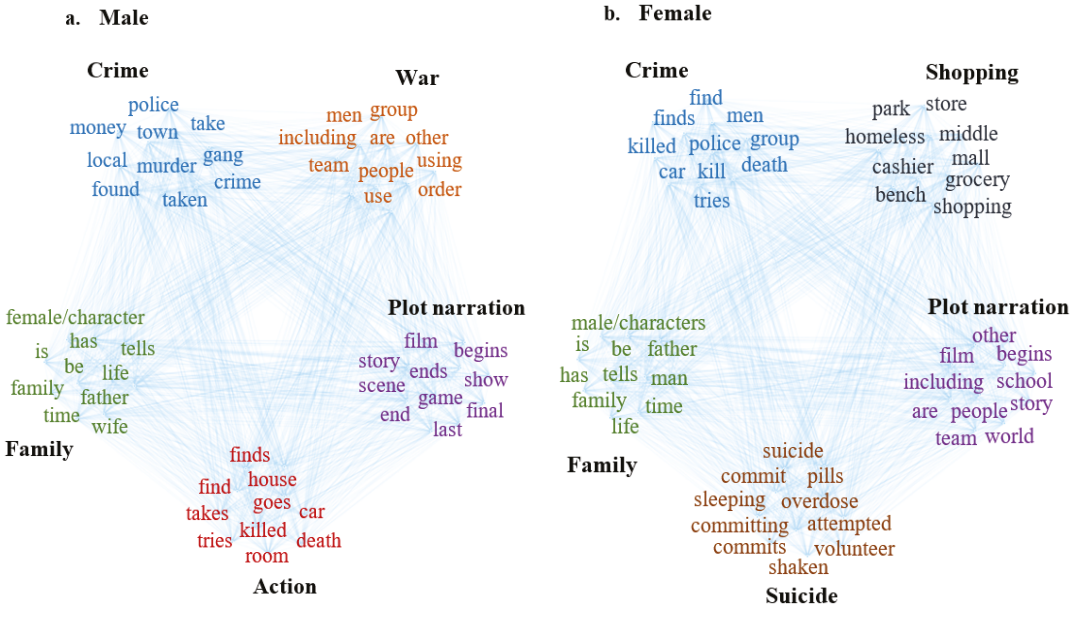


Figure 3. Communities identified in the co-occurrence networks. The male network (**a**) has 5355 vertices (words) and 1,979,829 edges (pairwise combinations of words within the data sample) and the female network (**b**) has 7393 vertices and 2,403,651 edges. We detected communities in the network using the Louvain algorithm. Five communities emerged in each of the networks, and the top ten vertices in terms of degree are shown.

3.2. Gender-Specific Story Tropes and Their Evolution

In this section, we describe the story tropes of male and female characters. We determined tropes by identifying significant paths in the network described by the path: 'character vertex–primary association–secondary association' (see Figure 4 for a visualisation of tropes in the network). Each 3-tuple of vertices denoted a significant story trope associated with each gender and their primary associations. To identify the most significant story tropes, we computed the path weights of all story tropes and selected the most significant tropes associated with each primary vertex. The top twenty most significant paths for male and female characters can be seen in Figure 5. The major difference between the tropes of male and female characters is that the paths of female characters are dominated by tropes that describe romance and family, whereas, for male characters, the most common tropes include friendship, family, romance, career, and crime/violence. The most significant trope for female characters is described by the path 'female character–love–fall'. For male characters, the most significant trope is described by the path 'male character–old–friend'.

Changes in stereotypical tropes across decades: We wanted to understand how stereotypical story tropes changed across time. First, we chose the most significant story trope from the two identified stereotypical domains (romance and crime/violence) and measured their path weights across each decade, from 1940 to 2019. For male characters, the most significant crime/violence-related trope was 'male–kill–attempts'. For female characters, the most significant romance-related trope was 'female–love–falls'. There were no crime/violence-related tropes in the female network.

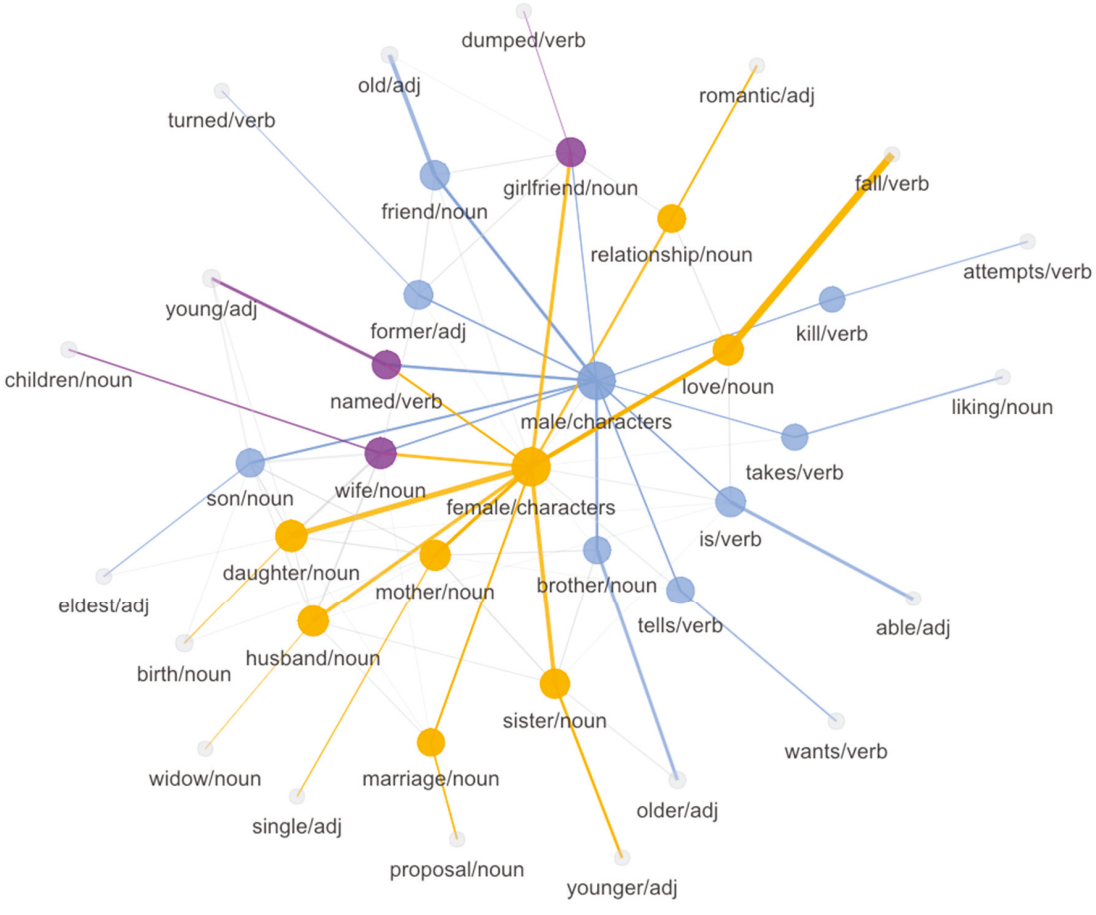


Figure 4. Most significant story tropes associated with male and female characters described by the path 'character–primary vertex–secondary vertex'. The thickness of the line represents the significance of the log-likelihood ratio. Blue nodes and lines represent unique associations with male characters; orange nodes represent unique associations with female characters. Purple nodes and lines represent the common associations of both male and female characters.

While the trope of male characters being associated with 'attempts to kill' increased in significance (R^2 = 0.55, beta = 0.40, p = 0.03), the association of male characters with the 'dumps girlfriend' trope was inconclusive. This was because 'dumped' was absent as a secondary co-occurrence of male characters in some of the decades, making the log-likelihood test and comparison across decades impossible. For female characters, the association with the 'falling in love' trope decreased in significance R^2 = 0.73, beta = −3.89, p = 0.01) (Table 2; see also Supplementary Materials for figures showing how the path weights of these tropes changed across time).

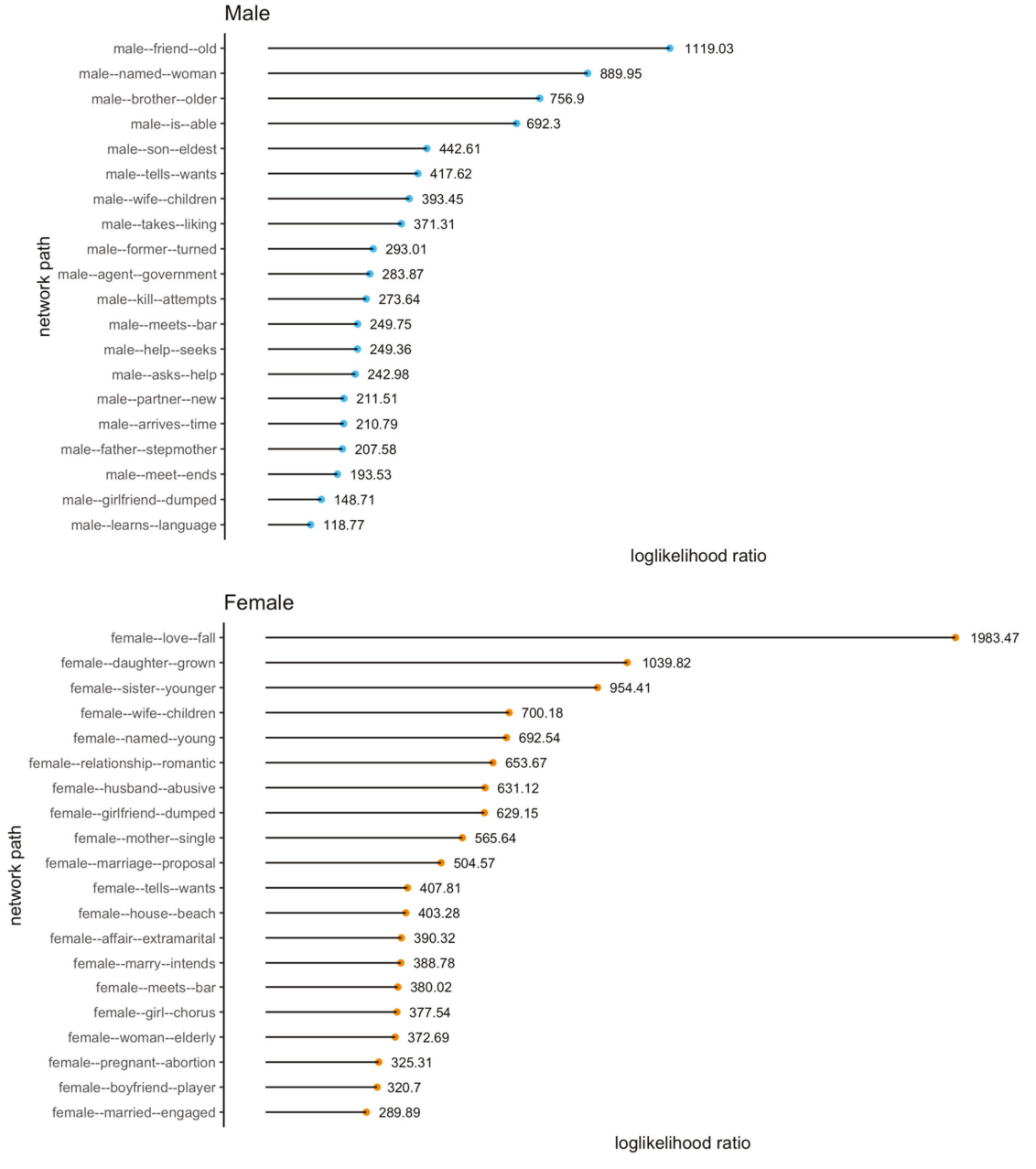


Figure 5. Needle plot representing the edge weights of the twenty most significant paths associated with male characters (**top**) and female characters (**bottom**).

Table 2. Linear regression analysis of path weight regressed on decade. * $p < 0.05$.

Male	R^2	Slope (Beta)	p
Romance			
Male-girlfriend-dumped		Inconclusive	
Crime/Violence			
Male-kill-attempts * [increasing]	0.55	0.40	0.03
Female	**R^2**	**Slope (Beta)**	**p**
Romance			
Female-fall-love * [decreasing]	0.73	−3.89	0.01
Crime/Violence			
nil			

In addition to this analysis, two networks were also created specifically for the 1940s and 2010s respectively to provide a visual representation of the changes in the significant character tropes (see Figure 6). There were several key differences between the networks of male and female characters between the 1940s and the 2010s. Firstly, there was an absence of marriage-related tropes in the network of the 2010s. There was a new trope linked to sexual relationships (female–relationship/noun–sexual/adj). However, the trope of falling in love was present in both networks. For male characters, there was a new trope linked to crime/violence (male–kill/verb–tries/verb). This trope was similar to the male–kill–attempts trope that was observed in the overall network (see Figures 4 and 5), which showed a significant increase across time.

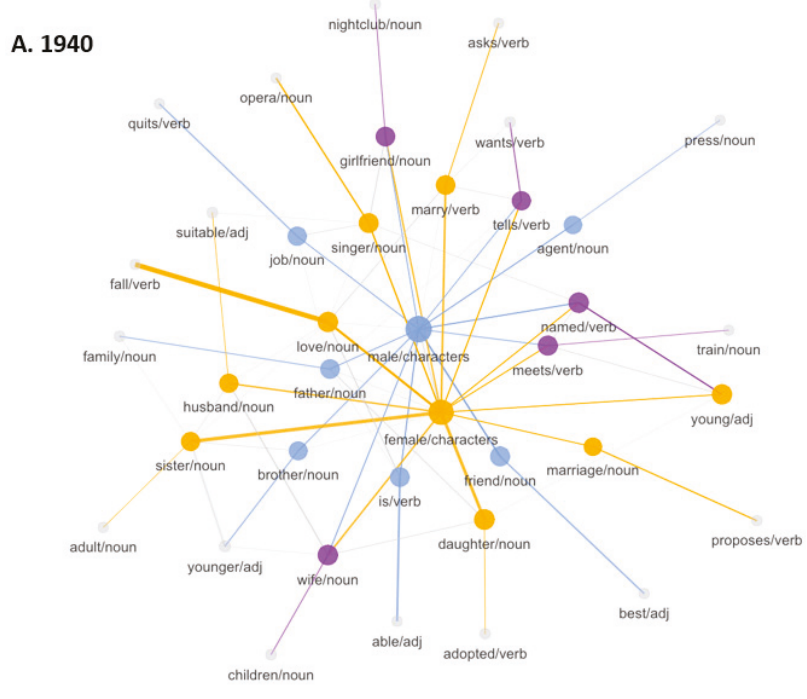


Figure 6. *Cont.*

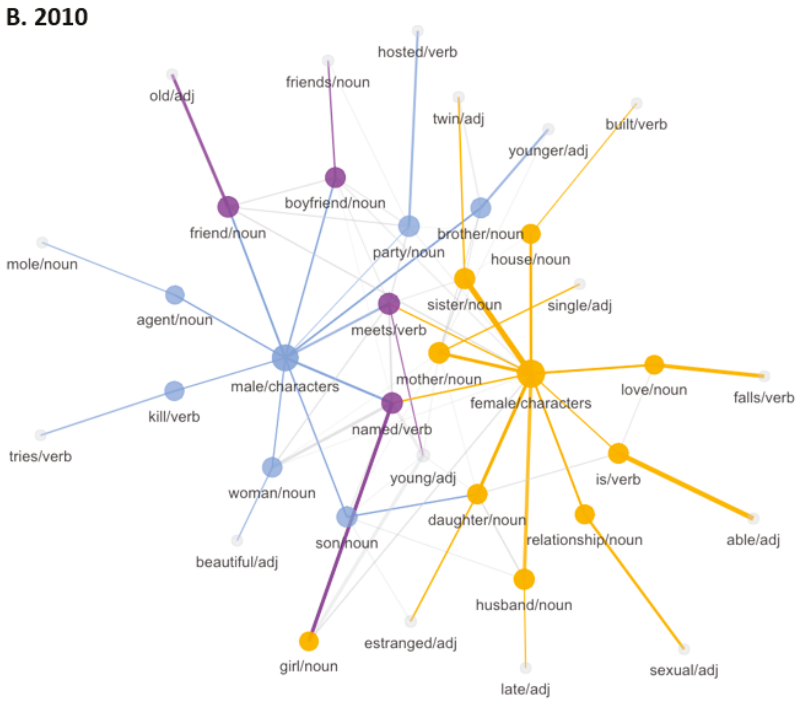


Figure 6. A comparison of significant tropes in the 1940s (**A**) and 2010s (**B**). Some key differences between the tropes from these two decades include the disappearance of the marriage trope in the female characters' network between the 1940s and the 2010s, and the new addition of a trope related to sexual relationships. For male characters, while there were no crime-related tropes in the 1940s, there was one in the 2010s.

3.3. Roles, Actions, and Descriptions of Male and Female Characters

3.3.1. Nouns: What Roles Do Male and Female Characters Play?

To understand the most common roles that male and female characters play in Hollywood movies, noun vertices with the highest edge weights in associated with the male and female character vertices were identified. A network with twenty of the most significant associations of each gender is presented in Figure 7. The edge weights for each of these associations are shown in Figure 8. It is evident from Figure 7 that both male and female characters played the role of 'blood relatives' in a family.

In addition, we identified individual nouns from both male and female character networks that fit the domains of crime/violence (e.g., 'murder') and romance (e.g., 'love'). We then fitted a linear regression model to investigate the changes in their edge weights across time. The results are listed in Table 3 (additional figures depicting the change in edge weights can be found in the Supplementary Materials). For male characters, none of the associations in the domain of crime showed significant changes across time. However, in relation to the domain of romance, the association of male characters with the word 'wife' decreased significantly in the same period. For female characters, the associations with 'love', 'wife', and 'widow' in the romance domain showed a significant decline across time, but the associations with 'relationship' and 'wedding' showed a significant increase across time. There were no nouns found in the female character network that were associated with crime.

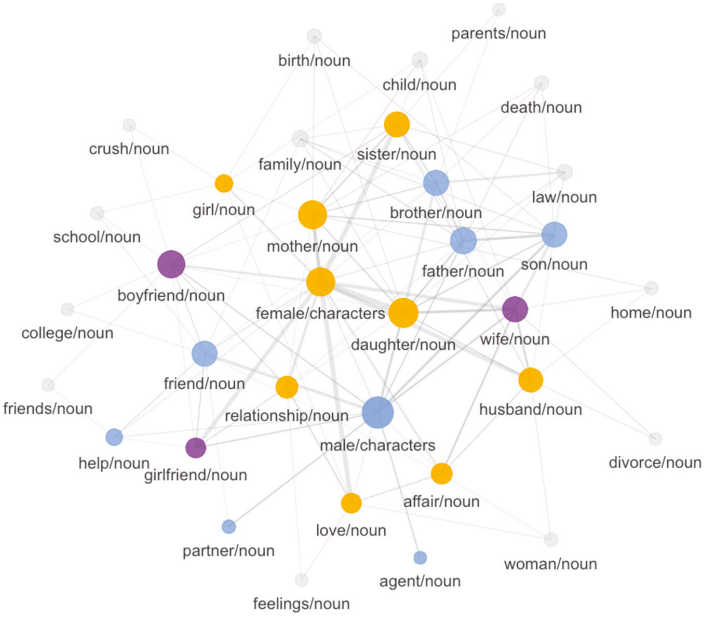


Figure 7. Most significant primary noun associations with males and females. Nouns found to strongly co-occur with female characters included 'daughter' and 'mother'; for male characters, they included 'friend' and 'father'. Nouns such as 'boyfriend', 'wife', and 'girlfriend' co-occurred strongly with both female and male characters.

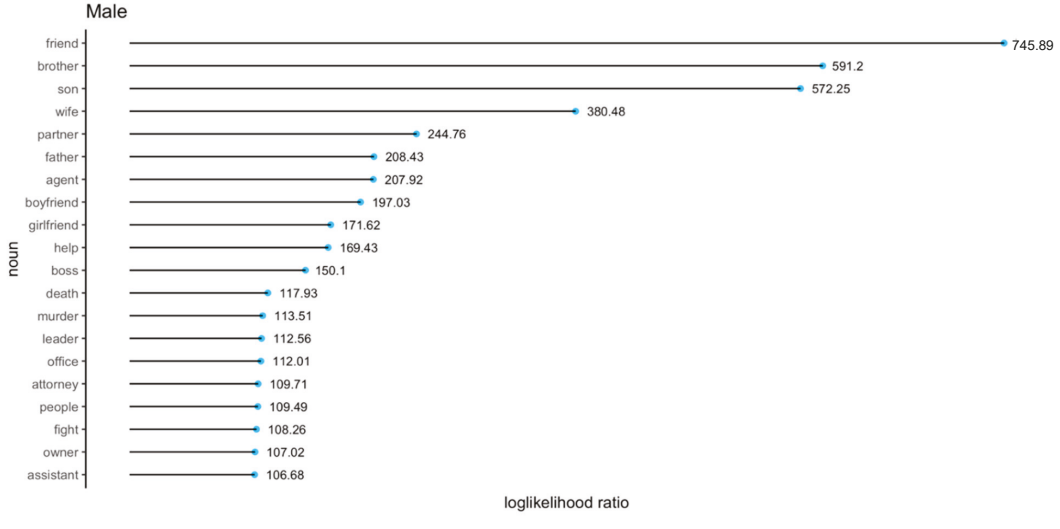


Figure 8. *Cont.*

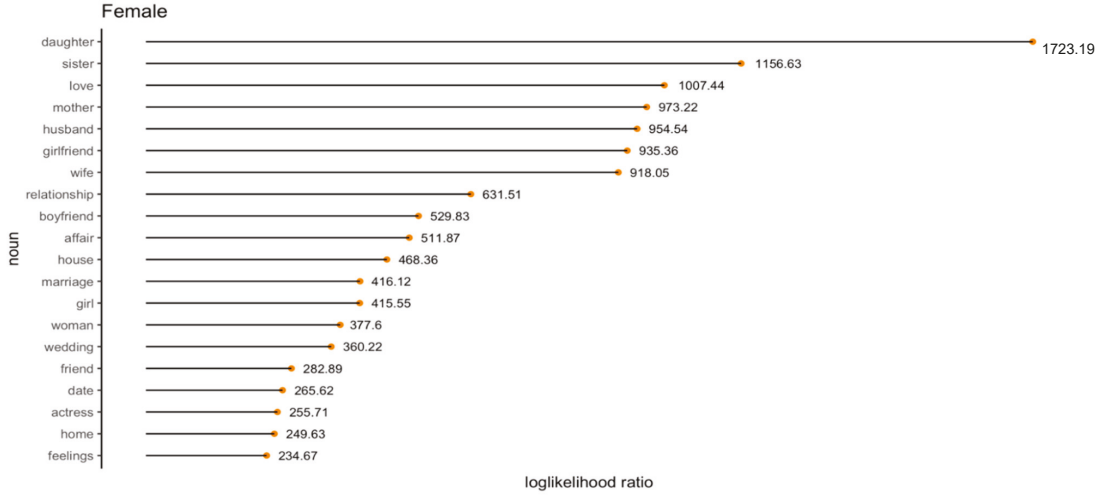


Figure 8. Needle plot representing edge weights of the twenty most significant primary noun associations of male characters (**top**) and female characters (**bottom**).

Table 3. Linear regression analysis of edge weight regressed on decade for stereotypical nouns. * $p < 0.05$, + $p < 0.10$.

Male	R^2	Slope (Beta)	*p*
Romance			
Wife * [decreasing]	0.58	−0.62	0.03
Girlfriend	0.14	−0.09	0.36
Boyfriend	0	0.02	0.91
Crime/Violence			
Death	0.15	0.11	0.34
Murder	0.27	−0.14	0.19
Gun	0.26	−0.14	0.19
Female	**R^2**	**Slope (Beta)**	***p***
Romance			
Love * [decreasing]	0.59	−1.03	0.03
Girlfriend	<0.01	0.03	0.95
Wife *	0.67	−1.15	0.01
Relationship + [increasing]	0.43	0.77	0.08
Affair	0.03	0.16	0.69
Marriage	0.37	−0.41	0.11
Wedding + [increasing]	0.45	0.29	0.07
Widow * [decreasing]	0.63	−0.44	0.02
Crush	<0.01	0.03	0.87
Crime/Violence			
nil			

3.3.2. Verbs: What Do Male and Female Characters Do?

To understand the most common actions that male and female characters perform in Hollywood movies, verb vertices with the highest edge weights associated with the male and female gender vertices were identified. A network with twenty of the most significant associations of each gender is presented in Figure 9. The edge weights for each of these

associations are shown in Figure 10. We identified several stereotypical verb associations for each gender based on the domains of crime/violence and romance and fitted a linear regression model to investigate the changes in their edge weights across time. When a word with multiple tenses was identified (e.g., 'marry', 'married'), we used the tense with the largest edge weight for our analysis. These are listed in Table 4 (additional figures depicting the change in edge weights can be found in the Supplementary Materials). For male characters, 'kill' (in the domain of crime) showed a significant increase across time. No verbs associated with romance were found in the male network. For female characters, the verb 'marry' (in the domain of romance) showed a significant decrease across time. No verbs associated with crime were found in the female character network.

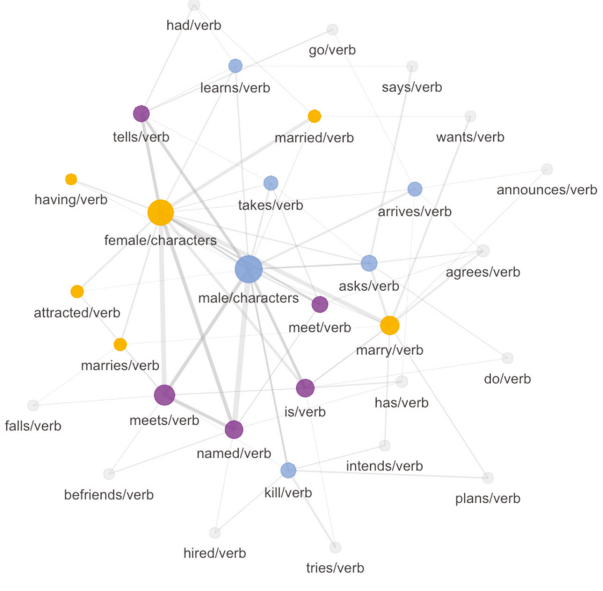


Figure 9. Most significant primary verb associations with males and females. Verbs found to strongly co-occur with female characters included 'marry' and 'married'; for male characters, they included 'kill' and 'arrives'. Verbs such as 'named' and 'meets' co-occurred strongly with both female and male characters.

Table 4. Linear regression analysis of edge weight regressed on decade for stereotypical verbs. ** $p < 0.01$, * $p < 0.05$.

Male	R^2	Slope (Beta)	*p*
Romance			
nil			
Crime/Violence			
Kill * [increasing]	0.55	0.17	0.03
Female	**R^2**	**Slope (Beta)**	***p***
Romance			
Marry ** [decreasing]	0.89	−1.15	<0.01
Attracted	0.31	−0.32	0.15
Loves	0.01	0.04	0.84
Dating	0.01	0.03	0.81
Crime/Violence			
nil			

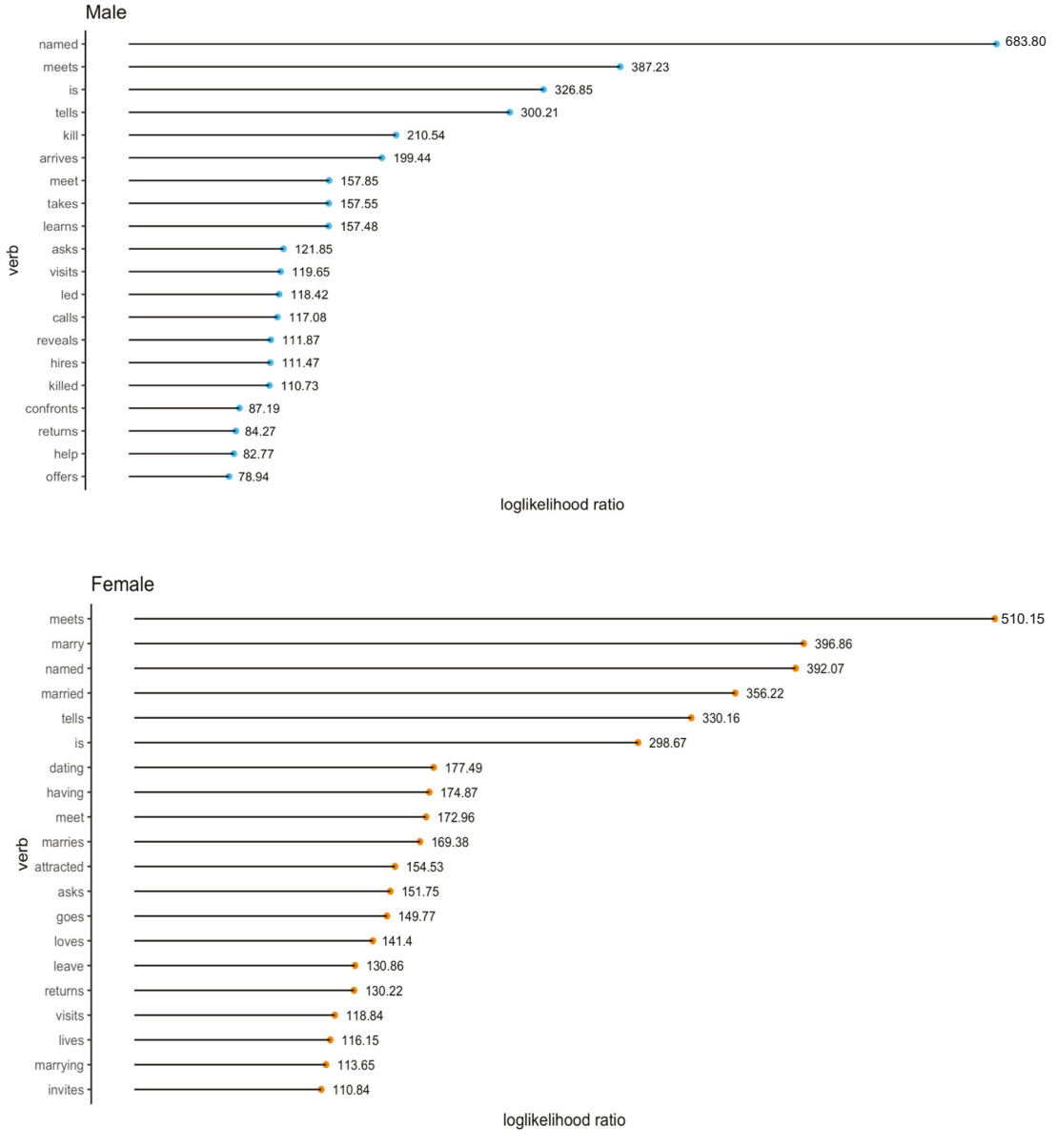


Figure 10. Needle plot representing edge weights of the twenty most significant primary verb associations of male characters (**top**) and female characters (**bottom**).

3.3.3. Adjectives: How Are Male and Female Characters Described?

To understand the most common descriptions of male and female characters in Hollywood movies, adjective vertices with the highest edge weights associated with male and female characters were identified. A network with twenty of the most significant associations of each gender is presented in Figure 11. The edge weights of these associations are shown in Figure 12. We identified stereotypical adjectives associated with violence/crime and romance for both male and female characters and fitted a linear regression model

to investigate their changes in edge weight across time. The results are listed in Table 5 (additional figures depicting the change in edge weights can be found in the Supplementary Materials). For male characters, 'corrupt' (related to the crime/violence domain) did not demonstrate a significant change across time. 'Married' (from the romance domain) did not demonstrate a significant trend either, whereas 'handsome' was inconclusive, as it did not co-occurred with male characters in some decades. For female characters, 'beautiful' and 'attractive' (romance) showed a significant decrease across time. No adjectives related to crime and violence were detected in the female character network.

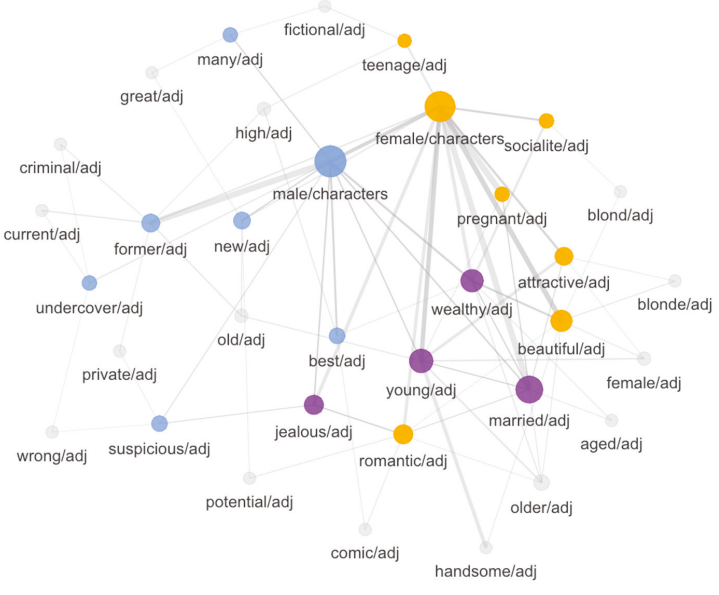


Figure 11. Most significant primary adjective associations with males and females. Adjectives found to strongly co-occur with female characters included 'pregnant' and 'beautiful'; for male characters, they included 'former' and 'best', among others. Adjectives such as 'young' and 'married' co-occurred strongly with both female and male characters.

Table 5. Linear regression analysis of edge weight regressed on decade for stereotypical adjectives. * $p < 0.05$.

Male	R^2	Slope (Beta)	*p*
	Romance		
Married	0.11	0.05	0.42
Handsome		Inconclusive	
	Crime/Violence		
Corrupt	0.04	−0.03	0.62
Female	**R^2**	**Slope (Beta)**	***p***
	Romance		
Beautiful * [decreasing]	0.70	−0.37	0.01
Attractive * [decreasing]	0.69	−0.32	0.02
Married	0.24	0.21	0.22
Romantic	0.24	−0.16	0.22
	Crime/Violence		
	nil		

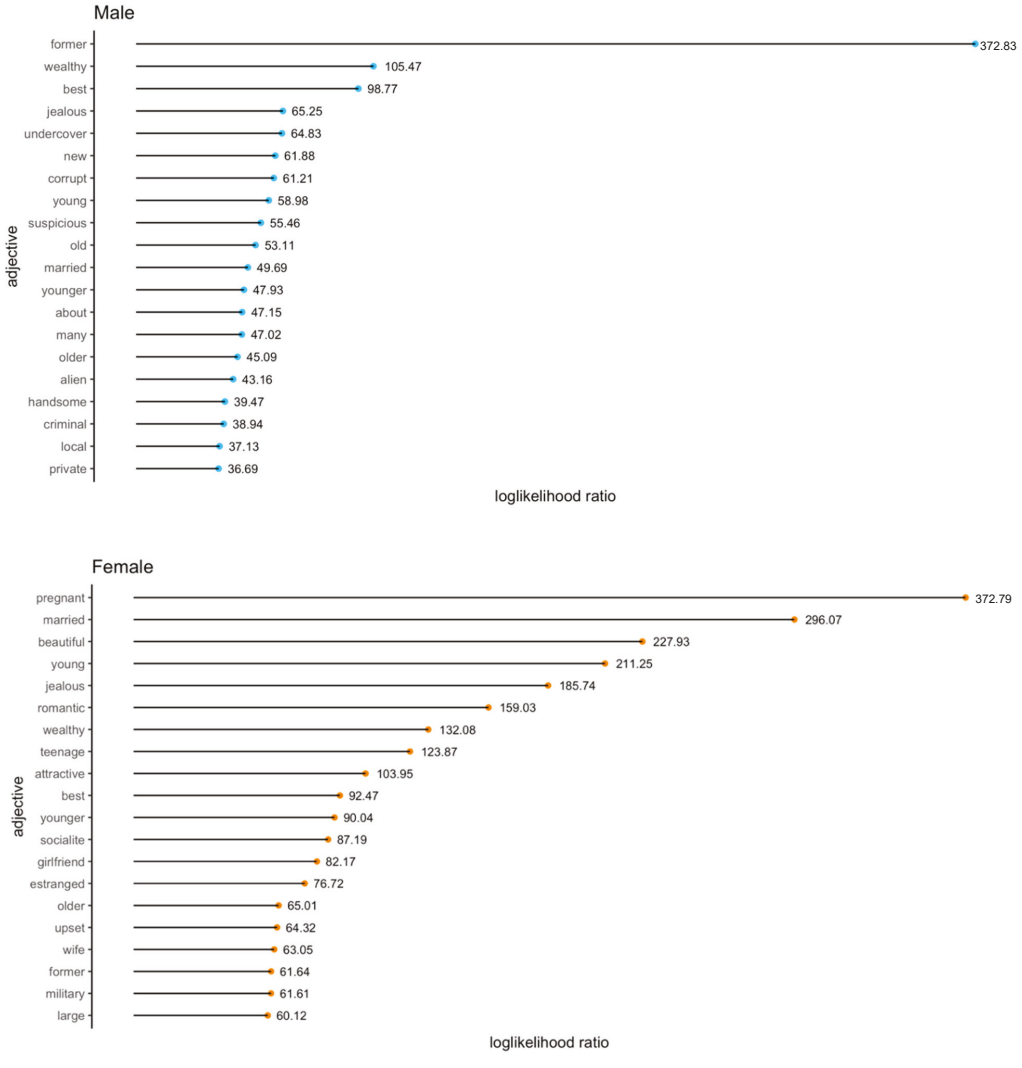


Figure 12. Needle plot representing edge weights of the twenty most significant primary adjective associations of male characters (**top**) and female characters (**bottom**).

4. Discussion

In this study, we used network analysis to understand the stereotypes associated with male and female characters in Hollywood movies using movie plot data scraped from Wikipedia. We used three different types of network analysis at different levels of the network for the purposes of our investigation. Through a community detection analysis, we aimed to understand the structure of the network at the meso-level and uncover the themes used in the portrayal of men and women. Through our novel approach of understanding tropes using path analysis and the analysis of the edge weights of individual words, we aimed to more specifically understand the lives of male and female characters and how stereotypical representations change with time. Overall, this paper provides empirical evidence that gender stereotypes are expressed through the cultural products of a society. We further demonstrate how the use of network analysis could be a compelling approach

to model the representations of different social groups in the products of society and study their dynamic changes over time. An overview of our findings based on our community analysis, trope analysis, and the analysis of individual nouns, verbs, and adjectives, can be found in Table 6. In the following sections, we discuss the key results of our community analysis, trope analysis, and edge-weight analysis of individual words.

Table 6. Summary of quantitative results from different types of analysis conducted. ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$.

Analysis		Findings
Community	Male	1. Family (1465 vertices) 2. Action (1408 vertices) 3. War (1405 vertices) 4. Plot narration (1405 vertices) 5. Crime (524 vertices)
	Female	1. Crime (2876 vertices) 2. Family (2547 vertices) 3. Plot narration (1941 vertices) 4. Shopping (17 vertices) 5. Suicide (12 vertices)
Story trope	Male (top 20)	male–friend–old, male–named–woman, male–brother–older, male–is–able, male-son–eldest, male–tells–wants, male–wife–children, male–takes–liking, male–former–turned, male–agent–government, male–kill–attempts, male–meets–bar, male–help-seeks, male–asks-help, male–partner–new, male–arrives–time, male–father–stepmother, male–meet–ends, male–girlfriend–dumped, male–learns–language
	Female (top 20)	female–love–fall, female–daughter–grown, female–sister–younger, female–wife–children, female–named–young, female–relationship–romantic, female–husband–abusive, female–girlfriend–dumped, female–mother–single, female–marriage–proposal, female–tells–wants, female–house–beach, female–affair–extramarital, female–marry–intends, female–meets–bar, female–girl–chorus, female–woman–elderly, female–pregnant–abortion, female–boyfriend–player, female–married–engaged
	Male (analysed)	**Crime/Violence** male–kill–attempts * [increasing] ($R^2 = 0.55$, beta = 0.40, $p = 0.03$) **Romance** male–girlfriend–dumped
	Female (analysed)	**Crime/Violence** nil **Romance** female–fall–love * [decreasing] ($R^2 = 0.73$, beta = −3.89, $p = 0.01$)
Noun	Male (top 20)	friend, brother, son, wife, partner, agent, father, girlfriend, boyfriend, help, boss, death, attorney, manager, owner, women, people, murder, gun, office
	Female (top 20)	daughter, sister, love, mother, husband, girlfriend, wife, relationship, boyfriend, affair, house, marriage, girl, woman, wedding, friend, date, actress, home, feelings
	Male (analysed)	**Crime/Violence** death, murder, gun **Romance** wife * [decreasing] ($R^2 = 0.58$, beta = −0.62, $p = 0.03$), girlfriend, boyfriend

Table 6. *Cont.*

Analysis		Findings
	Female (analysed)	**Crime/Violence** nil **Romance** love * [decreasing] (R^2 = 0.59, beta = −1.03, p = 0.03), girlfriend, wife * [decreasing] (R^2 = 0.67, beta = −1.15, p = 0.01), relationship + [increasing] (R^2 = 0.43, beta = 0.77, p = 0.08), affair, marriage, wedding + [increasing] (R^2 = 0.45, beta = 0.29, p = 0.07), crush, widow * [decreasing] (R^2 = 0.63, beta = −0.44, p = 0.02)
Verb	Male (top 20)	named, meets, tells, is, arrives, kill, takes, meet, learns, asks, killed, visits, hires, led, returns, kills, suspects, calls, convinces, sends
	Female (top 20)	meets, marry, named, married, tells, is, dating, having, meet, marries, attracted, asks, goes, loves, leave, returns, visits, lives, marrying, invites
	Male (analysed)	**Crime/Violence** kill * [increasing] (R^2 = 0.55, beta = 0.17, p = 0.03) **Romance** nil
	Female (analysed)	**Crime/Violence** nil **Romance** marry ** [decreasing] (R^2 = 0.89, beta = −1.15, p < 0.01), attracted, loves, dating
Adjective	Male (top 20)	former, wealthy, best, jealous, married, suspicious, young, new, undercover, many, old, real, older, younger, private, interested, local, corrupt, about, handsome
	Female (top 20)	pregnant, married, beautiful, young, jealous, romantic, wealthy, teenage, attractive, best, younger, socialite, girlfriend, estranged, older, upset, wife, former, military, large
	Male (analysed)	**Crime/Violence** corrupt **Romance** married, handsome
	Female (analysed)	**Crime/Violence** nil **Romance** beautiful * [decreasing] (R^2 = 0.70, beta = −0.37, p = 0.01), attractive * [decreasing] (R^2 = 0.69, beta = −0.32, p = 0.02), married, romantic

4.1. Community Analysis

First, we conducted a community analysis of a co-occurrence network based on the words associated with male and female characters to understand the themes associated with these characters. The results of our community analysis have to be interpreted with caution because of the low modularity values, which indicate lack of robust community structure. To the best of our knowledge, only Xu et al. [23] have performed a similar analysis of movies, albeit using movie synopses from IMDb. They identified communities of crime, action, career, and family in association with male characters and the communities of action, romance, career, and family in association with female characters. By contrast, our analysis showed that male characters were associated with the themes of family, crime, action, and war, whereas female characters were associated with family, crime, shopping, and suicide. One possible explanation for these differences is that our present analysis considered both lead and supporting characters, whereas Xu et al. [23] only considered lead characters in movies.

The results of our analysis partly support social role theory, which states that the stereotypes used in a society reflect the gender roles and expectations of that society. For

example, the specific association of war with male characters could be partly attributed to the large number of movies related to World War II in Hollywood [37]. War was often used as a central theme in movies, depicting major social issues such as the post-war readjustment to life of male soldiers. On the other hand, the association of the theme of 'suicide' with female characters seems particularly noteworthy. In reality, suicide rates among men are higher than among women, with the US reporting 22.8 male suicides versus 6.2 female suicides per 100,000 people in 2018—a disparity that has remained relatively stable for the past 60 years [38]. It would thus be expected, according to social role theory, that men would be more commonly associated with suicide movies than women in movies, which is contrary to our results. Specifically, suicide is often negatively constructed and stigmatised by many. Often resulting from the need to escape from psychological pain [39], the act of suicide and its ideation are used to emphasise the vulnerability and incompetence of female characters [40]. However, it is worth noting that although men are more likely to commit suicide than women, the trend is reversed for attempted suicides—women are more likely to attempt suicide than men [41]. It is possible that the significant portrayal of suicide in relation to female characters, but not male characters, in movies reflects the general perceptions (including misconceptions) people hold about suicide (i.e., that it is more associated with women than men). It was also surprising to observe that female characters were associated with the theme of crime, as this was primarily a stereotype of male characters in the media, according to our initial literature review.

4.2. Trope Analysis

Next, we identified story tropes in the networks of male and female characters using a novel method of path analysis implemented on the word co-occurrence networks formed using the significance of log-likelihood ratios. Story tropes add a further dimension to the standard single-word associations that have often been used to study stereotypes. Our results showed that the lives of female characters are dominated by the tropes of romance and family. By contrast, the lives of male characters are filled with more diverse aspects of life, including friendship, career, and crime/violence, as well as family and romance. This is not surprising given that the portrayal of female characters as passive love interests has been present in Hollywood movies for a long time, whereas male characters "get in on the action" [42]. To allow gender-based comparisons across time, we focused on crime/violence and romance domains.

Our analysis of the changes in the path weights of the most significant stereotypical tropes for both male and female characters showed that the trope of a male character attempting to kill increased. For females, the trope of falling in love has decreased, and no significant crime/violence-related tropes were found in the network. Even though changes in a few specific tropes would be insufficient for us to conclude that stereotypical associations have changed overall, these results do support the notion that some of the common and stereotypical associations of males and females have in fact changed in movies. Our edge-weight analysis of individual words, discussed below, offers a deeper understanding of the nature of these changes. Specifically, although we saw a general increase in the crime/violence trope in association with male characters, and a decrease in the romance trope in association with female characters, the trends in the associations of individual words within those domains may not necessarily be the same. Within the romance domain, for instance, we could explore trends in discrete words specifically related to marriage, or courtship separately, which portray a richer picture of what comprises these stereotypes.

4.3. Edge-Weight Analysis

Lastly, we studied the edge weights of the most significant crime-related and romance-related noun, verb, and adjective associations of male and female characters. From the overall analysis across the entire time period, we found several interesting trends in how stereotypical domains within the genders evolved over time. For male characters, most

of the words related to crime (i.e., 'death', 'murder', 'gun', 'corrupt') did not show any significant change in co-occurrence from the 1940s to the 2010s, except for the specific verb, 'kills', which was increasingly associated with male characters in Hollywood movies during this period. For romance-related words, on the other hand, only 'wife' demonstrated a significant decrease across time in association with male characters. The words 'girlfriend', 'boyfriend', and 'married' displayed no discernable trends. For female characters, by contrast, we found a mix of increasing, decreasing, and unchanging trends in various romance-related words. The co-occurrence of words such as 'love', 'wife', 'widow', 'marry', 'beautiful', and 'attractive' with female characters decreased over the years, whereas 'relationship' and 'wedding' increased. The romance words that did not display a significant change in co-occurrence with female characters during this period included 'affair', 'marriage', 'crush', 'girlfriend', 'attracted', 'love', 'dating', 'married', and 'romantic'. Interestingly, no crime-related words were detected in the female character networks. Furthermore, there was a general lack of significant romance-related noun, verb, and adjective associations for male characters and a lack of significant associations for female characters in the crime domain. In some instances, no relevant words could be detected and thus analysed (e.g., romance-related nouns for male characters, crime/violence-related nouns, verbs, and adjectives for female characters).

Our edge-weight analysis of the noun, verb, and adjective associations showed that the representation of male characters through crime-related words generally increased over the decades. The analysed words either remained highly associated with male characters or increased significantly, further perpetuating the idea that Hollywood movies continue to portray male characters in relation to crime and violence. This reinforces male characters' disposition to partake in crimes that could potentially involve violence or some form of dishonesty as a way to display masculine identities and power [43]. Participation in crime is believed to reflect hegemonic masculinity, according to which males regard engagement in crime as an outlet for their aggression, reinforcing their power and prowess [44].

This long-standing stereotype associated with men, crime, also stems from conventions of masculine strength. For example, the act of stalking, which is typically depicted by the media as a gendered crime, involves the "popular image . . . . . of a celebrity who is stalked by a crazed fan or a battered woman who has left a physically abusive relationship and is now being stalked by her ex-spouse or ex-lover" [45,46]. Unsurprisingly, in the movie entries that we analysed, there also seemed to be a significant link between the portrayal of men and criminally offensive roles [47]. The aggressive portrayal of men and its contrast with the depictions of women as helpless are also in line with the lack of crime/violence-related words found in the female characters' noun, verb, and adjective networks. From these findings, it seems that crime remains much more central to the narratives of male characters than to that of female characters.

The depiction of romance in relation to male characters, on the other hand, is generally sparser than that of female characters, as suggested by the smaller number of romance-related nouns, verbs, and adjectives found in the male character network compared with the female character network. However, the significant decrease in the association of male characters with the word 'wife' is interesting, possibly suggesting that male characters may be less frequently portrayed as married than they were previously. This could reflect a larger societal trend, in which marriage in the U.S. is declining [48]; however, given the lack of change observed in the use of the adjective 'married', this conclusion may be premature. In sum, it seems that romance is not a significant part of the lives of male characters in Hollywood movies as compared with female characters.

However, the results for female characters show a more ambiguous picture. Specific words related to marriage demonstrated both increasing (e.g., 'wedding'), decreasing (e.g., 'marry'), and unchanging (e.g., 'marriage', 'married') trends in relation to female characters in Hollywood movies across time. Traditionally, the role of woman has centred around marriage and the family [49]. In the 1950s, American culture strongly emphasised the creation of nuclear families (or the 'All-American family'), which reinforced the normative

expectation that women should take on the roles of wife and mother [50]. As expected, these traditional ideas around women and marriage are also evident in movies, such as the early Disney Princess movies, *The Little Mermaid* (1989) and *Aladdin* (1992), and even *Tangled* (2010), all of which conclude their female protagonists' narratives through a romantic, royal marriage, despite the fact that the premise of their individual stories is a search for autonomy and the pursuit of their dreams [51]. Evidently, women and matrimony are strongly associated in both American society and American movies, at least historically.

More recently, some researchers have observed or predicted a waning trend in these stereotypes of marriage. Barber [52] recognised that modern Disney Princess movies have shifted towards starring an independent female lead who is not bound by marriage. Examples include *Brave* (2012), *Frozen* (2014), and *Moana* (2016), all of which feature female protagonists who either actively reject marriage (e.g., Merida in *Brave*) or do not have a romantic arc that culminates in a marriage in their story at all (e.g., Elsa in *Frozen*, Moana in *Moana*). Such changes may be related to the decline in marriage in American society—marriage rates in the US have been steadily falling over the past decades, with around only 50% of American adults being married in 2016 compared to 72% in 1960 [48]. Indeed, the gendered expectation that a woman should marry was largely driven during a time when women were systematically disabled economically and marriage was a means of financial advantage and survival [53]. Such concerns may no longer be as relevant today as American women have higher earning power than before. However, our edge-weight analyses did not support the predictions that marriage would be less frequently depicted in female characters' arcs in more recent movies, as the majority of marriage-related words either remained significantly high or increased in association with female characters between the 1940s and 2010s. This suggests that overall, marriage continues to be a stereotypical trope in the lives of Hollywood's female characters, which is not reflective of trends in social reality.

Apart from matrimony, we also looked at words that emphasise women's romantic roles in general, such as 'wife', 'widow', 'crush', and 'girlfriend'. These nouns describe female characters in relation to their romantic partners, and have either decreased (e.g., 'wife', 'widow') or not changed (e.g., 'crush', 'girlfriend') in their association over time. The decreasing trends could point towards the increasing complexity of female characters, moving beyond the roles of 'wife' or 'widow'. This is in line with past research, which identified that female characters are represented with more complexity, as demonstrated through the increasing representation of the themes of 'self-interest' and 'protection of others' in their narratives in recent times, as compared to the conventional exclusive focus on romantic roles [54]. Perhaps female characters are now being represented through arcs that incorporate multiple social roles, reducing the focus on their relations to their husbands. However, viewed in the context of the lack of change in the associations of 'crush' and 'girlfriend', it is possible that only marriage-related roles are gradually being dropped from female characters' arcs, whereas courtship, or romance roles in general, continue to be represented in movies. Indeed, as marriage numbers have fallen in the U.S., there has been a rise in cohabitation among unmarried couples [55]. Critically, marriage and romance are not synonymous. These results could reflect real-life modernising attitudes and behaviours, in which individuals continue to pursue romance without symbolising it through marriage.

Lastly, an interesting result was the decrease in the associations of the words 'beautiful' and 'attractive' with female characters, which may suggest that women in Hollywood movies are now described through their physical appearance less than they were previously. The depiction of women in the media has long been criticised for pandering to the male gaze, which makes women "a passive object to be looked at" [56]. Accordingly, the camera is likened to the male audience, with its pans and angles over women's bodies and faces representative of the male desire to view women's physicality (which is often related to women's sexuality). In our edge-weight analysis, we found that female characters in Hollywood movies have, in fact, become less likely to be described through their beauty or attractiveness, which may signify a declining trend in the representation of women

in order to fulfil the "male gaze". Indeed, feminists have long fought against the female consumption of the beauty industry, which is driven, as suggested, by a forced competition for male approval [57]. Possibly, the shift towards a less beauty-oriented stereotypical representation of female characters in Hollywood reflects the increasing recognition or acceptance of these concerns.

4.4. Summary of Results

Summing up, it is clear from the results of our analysis that men and women are represented differently in cinema. We did not lend much weight to the results of the community analysis due to the low modularity value of the network's community structure. However, based on the other two sets of analyses, our results show that male and female characters are associated with several stereotypes that are commonly observed in real life, and which was in accordance with social role theory. For example, the lives of female characters in movies mainly revolve around romance and relationships, although this is not true for male characters. The roles, actions, and descriptions of female characters are also centred around romance and relationships. By contrast, the lives of male characters are much more diverse. There are aspects of career and crime/violence, but also romance and family, although romance was observed to be less of a relevant domain to male than to female characters. The roles, actions, and descriptions of male characters also reflect this diversity. While we identified both romance- and crime/violence-related words in the male networks, the female networks displayed an absence of crime/violence words.

When comparing networks of word co-occurrences across the entire time period from the 1940s to the 2010s, we observed several differences between the lives of male characters and the lives of female characters. In sum, our results seem to suggest that the stereotypical depiction of an 'attractive' and 'beautiful' female character who passively 'falls in love with a male character and marries him' has decreased. The trope of a female character as an 'attractive' and 'beautiful', but also a passive, character who 'falls in love with a male character and marries him' has decreased. This was inferred from the falling association of female characters with the trope of 'falling in love' and also with words such as 'attractive', 'beautiful', 'marry', 'love', and 'widow'. However, this does not quite change the overall stereotypical association of female characters with romance because even as these tropes and associations are in decline, they seem to be replaced by others. For example, associations with 'wedding' and 'relationship' appear to have increased. We also noticed new tropes that link females to sexual relationships, despite the lack of tropes linked to marriage in the 2010s network. Due to the mixture of trends observed, a more suitable conclusion may be that the ways in which female characters are represented through romance are complex and continually evolving, rather than purely declaring a quantitative rise or fall. As for male characters, the associations seem to be largely stable; however, association with crime seems to have increased in some cases, as inferred from the rising association with the trope of 'male as a killer' and the word 'kill'.

According to social role theory [10], stereotypes mirror societal roles and expectations. Through our examinations of gender stereotypes in Hollywood movies, we showed that in some cases, this is indeed the case, whereas it is not necessarily true in other cases. Therefore, it appears that the stereotypes of a society, as represented in its cultural products, need not actually represent gendered roles and expectations in social reality. That is, social role theory might not always apply to the implicit stereotypes of a society as manifested in its cultural products. An alternative interpretation could be that stereotypes are not absolute categorisations. What is considered a stereotype could consist of a plethora of associations, with each association changing at a different pace. Therefore, although it is helpful to group together specific associations under a single broad categorisation, specific associations should be studied to gain a nuanced understanding of how stereotypes change within society. It may be more meaningful to view gender representations as being subject to constant development and change, and leverage the availability of big

data and computational and quantitative methods to obtain nuanced insights into specific stereotypical associations.

4.5. Real-Life Implications

The harmful effects of stereotypes have been well-documented in psychology. For example, stigmatised populations perform worse when they recognise the negative stereotypes surrounding them and their abilities, such as women and mathematical skills, referred to as the stereotype threat [58]. These effects have been observed in female gamers and women in engineering [59,60]. Furthermore, gender stereotypes in children's books have been found to affect children's cognitive development, potentially negatively skewing their world-views [61]. In light of these findings, we should seek ways to reduce the perpetuation of gender stereotypes, in which the media is largely responsible for perpetuating.

The idea that audiences learn and model content from the media is not new, as explained through social learning theory [62]. Indeed, research has found that gender stereotypes presented in video games (e.g., the aggressive male character versus the sexualised, objectified female character) are mirrored in the perceptions of youth [63]. Exposure to stereotype-confirming information, such as the association between Arab people and terrorist acts, has also been found to prime negative perceptions of the stereotyped group among audiences [64]. Evidently, the media has deep and far-reaching implications for the ways in which people understand, perceive, and behave in the world. Long-term exposure to gender-stereotypical content in the media may potentially result in distorted beliefs about genders. For example, repeated exposure to stereotypically aggressive and criminal male characters may shape unbalanced views of the virtues of the respective genders (e.g., the 'women-are-wonderful' effect; [65]). Repeated exposure to stereotypically romance-oriented female characters may implicitly teach girls what to prioritise in their lives. Based on our study, we broadly confirm that Hollywood movies are one form of media in which gender stereotypes are presented, and there are diverse trends in more specific stereotypes as they rise or fall across time. This knowledge can prompt mass audiences to be more critical of the content they consume, and caution against accepting the stereotypical representations of male and female characters as congruent with reality. For filmmakers, these findings can guide them to be more aware of the ways in which they characterise men and women in cinema, and steer them away from perpetuating damaging stereotypes.

Finally, the framework we used in this study may be of use to researchers interested in documenting and examining stereotypes in other text-based media modalities or forms (e.g., movie scripts, television subtitles, books, and newspapers). There is already an existing evidence base suggesting negative stereotyping of multiple, diverse groups of people in the media, including British Muslims [66] and older adults [67]. As such, research on stereotypes presented through cultural products that was previously conducted with relatively small samples and through researcher-coded content analysis methods may benefit from our network approach, which allows large-scale data to be included in the analysis to obtain more widespread, systematic findings.

5. Limitations

Despite the theoretical, empirical, and methodological advances of our study, it is important to reflect on the limitations of our study as well. The main limitation is that we analysed North American movies, owing to the ease of availability of data on Hollywood movies. We also wished to leverage and build on the vast existing literature on North American society. In addition, a lack of non-binary characters from movies across the decades resulted in a lack of representation of these characters in our analysis. This lack of representation of non-binary characters across decades could be attributed to recent changes in our society; the identification of non-binary genders has only increased over the past few years [68]. The genderizeR package that was used in our analysis of characters and their genders also lacked the data to code non-binary characters. Hence, our results may

not be easily generalised to societies and movies from around the world, especially given the lack of transnormativity in the movies in general. Future studies of movie plots could be conducted to compare the nature of gender representations across different regions (e.g., movie plots from Bollywood or globally recognised movie industries, such as the cinema of Hong Kong). This could then be correlated with indexes on gender equality in the region. Such an analysis would add another level of investigation to the question of whether a society's gender equality status has an influence on the representation of gender in its cultural products.

Second, we were not able to capture words that were semantically similar to each other to understand the changes in the associations of entire concepts. This could be made possible by applying other methods, such as textual forma mentis networks [19]. Future studies could use this approach in addition to the approach we employed here.

Third, we have to acknowledge and accept a certain level of ambiguity when it comes to interpreting stereotypes from word networks. Due to the deconstructive nature of a network, which segments sentences via smaller semantic units such as words, there is an inevitable loss of the context that is expressed through the original, full sentence. As such, given an association such as '[male name]-killed', we cannot confirm if the associated meaning is being 'acted on' or 'received by' the character (i.e., did he kill someone, or was he killed?). Hence, we must recognise that underlying assumptions were made when interpreting the word networks. A possible way to counter this issue in future studies is to extend beyond employing the 3-tuples that were used in the trope analysis. As more words or data are captured in them than single-word associations, they may reflect more of the contextual information behind these specific associations, allowing us to draw more accurate conclusions as to the relationship between a character and their associated words.

Fourth, the 'genderizeR' package uses a constantly updating profile of social media profiles and their associated genders to tag first names as male or female. A shortcoming of this is that the names in the early part of the time period might have been less accurately tagged with gender information, despite the high accuracy of the package's identification of recent names. This is because the older demographic are less likely to use social media than younger segments of the population. Additionally, our entries of movie plots came from Wikipedia. While changelogs of these movie entries are available on Wikipedia, which allows the public to view page histories and make necessary corrections, we acknowledge that given its open-source nature, Wikipedia entries could still be relatively unreliable. Further studies could delve into using sources such as IMDb, where public contributions have to be reviewed by an editing team before they are published [69]. An additional standardised review process by a professional team would ensure that the information posted upholds a certain standard of reliability and accuracy. It is also important to acknowledge that given the reconstructed nature of Wikipedia plots, they are not to be taken as precise or exact representations of movies. In other words, the plot as written by a viewer (or multiple viewers, or editors, on Wikipedia) is an agreed-upon interpretation of the movie from which it is taken, independently of the movie itself. Hence, one must be cautious in interpreting the results of our analyses as completely accurate depictions of stereotypes in movies.

Lastly, to make our analysis tractable, we limited our story trope analysis to a specific path type (i.e., character vertex–primary vertex–secondary vertex). However, other types of paths and features of the network could be analysed using a variety of network science methods to reveal other aspects of the lives of male and female characters.

6. Conclusions

Using network analysis, this research investigated the trajectories of various gender stereotypes in movies via word co-occurrences generated from Wikipedia entries of movie plot descriptions. While the literature on gender stereotypes in the media is quite extensive, less research on this topic has been performed using quantitative approaches. Our approach is unique in its application of network analysis to analyse word co-occurrence patterns in

movie plots, coupled with a detailed longitudinal analysis of how such stereotypes evolve and change with time. This approach allowed us to explore gender stereotypes in the media and in society. We explored the idea that movies, as cultural artefacts, may reflect the gender stereotypes of society at large. However, due to the complex and dynamic nature of stereotypes, it seems that movies alone do not accurately capture the evolution of stereotypes in society. Given the dominance of the Hollywood movie industry, which continues to grow in terms of production and consumption (e.g., a record 873 movies were released in the US and Canada in 2018, compared with a mere 371 in 2000 [70]; box office revenue in the US and Canada exceeded 11 billion USD in 2019 versus a mere 1.66 billion USD in 1980 [71]), coupled with the potential of audiences to learn and model media content, developing an understanding of the types of stereotypes that either persist or wane in movies is important for studying how they may impact people's attitudes, emotions, and behaviour on a mass scale. Psychological research has long documented the harmful effects of stereotypes on stereotyped groups (e.g., [72]). Hence, our research provides one approach to the monitoring of the state of gender representation in Hollywood cinema, mapping its relation with changes in the values and norms of North American society. Through this, our approach provides new insights into specific gender stereotypes and their dynamic changes.

Supplementary Materials: The following document is available online at https://www.mdpi.com/article/10.3390/bdcc6020050/s1, Figure S1: Changes in path weights of stereotypical tropes across decades. (Left) Path weight of 'male–kill–attempts' (rep-resenting the crime/violence trope) significantly increased over time (R2 = 0.55, beta = 0.40, p = 0.03). (Right) Path weight of 'female–love–fall' (representing the romance trope) significantly decreased over time (R2 = 0.73, beta = −3.89, p = 0.01); Figure S2: Linear regression models inves-tigating changes in the edge weights of identified stereotypical primary noun associations of male and female characters. (a) The association of male characters with the noun 'wife' decreased sig-nificantly over the years. (b) Associations of male characters with nouns relating to crime did not significantly change across the years. (c) Associations of female characters with certain romance nouns, 'love', 'wife', and 'widow', significantly declined over the years, while associations with 'relationship' and 'wedding' significantly increased; Figure S3: Linear regression models inves-tigating changes in the edge weights of identified stereotypical primary verb associations of male and female characters. (a) The association of male characters with the verb 'kill' significantly in-creased over the years. (b) Associations of female characters with the verb 'marry' significantly fell over the years, whereas associations with 'attracted', 'loves', and 'dating' did not change signif-icantly; Figure S4: Linear regression models investigating changes in the edge weights of identi-fied stereotypical primary adjective associations of male and female characters. (a) The association of male characters with the adjective 'married' did not significantly change over the years. (b) The association of male characters with the adjective 'corrupt' did not significantly change over the years. (c) Associations of female characters with the adjectives 'beautiful' and 'attractive', relating to romance, significantly fell over the years, whereas associations with 'married' and 'romantic' did not change significantly.

Author Contributions: Conceptualisation, A.M.K., J.Y.Q.G. and T.H.H.T.; methodology, A.M.K., J.Y.Q.G. and T.H.H.T.; formal analysis, A.M.K., J.Y.Q.G. and T.H.H.T.; writing—original draft preparation, A.M.K., J.Y.Q.G. and T.H.H.T.; writing—review and editing, A.M.K., J.Y.Q.G., T.H.H.T. and C.S.Q.S.; Visualisation, A.M.K., J.Y.Q.G. and T.H.H.T.; supervision, C.S.Q.S.; funding acquisition, C.S.Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: C.S.Q.S. is supported by a Start Up Grant (WBS R-581-000-242-133) provided by the National University of Singapore.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data underlying the results presented in this study are available from the Wikipedia website (www.wikipedia.com, accessed on 12 April 2022). All Supplementary Materials, code, and processed data are available from https://osf.io/5pvj7 (accessed on 12 April 2022) and https://github.com/arjunkmrm/film-networks (accessed on 12 April 2022).

Acknowledgments: Portions of this research were completed as part of the course requirements for PL4246 Networks in Psychology, a capstone module taught at the Department of Psychology, National University of Singapore. The authors wish to acknowledge Russell Lee and the students of PL4246 for their comments and support.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

A1—Significance of log-likelihood ratio calculation
H_0: P(A | B) = P(A | −B)
H_1: P(A | B) ≠ P(A | −B)

$$\log \lambda(A, B) = -2.\log \frac{L(H_0)}{L(H_1)}$$

$$sig(A,\ B)_{lgl} = -2\log \lambda$$

$$\begin{aligned}\lambda = [n\log n- \quad & n_A \log n_A - n_B \log n_B + n_{AB} \log n_{AB} \\ & +(n - n_A - n_B + n_{AB}).\log(n - n_A - n_B + n_{AB}) \\ & +(n_A - n_{AB})\log(n_A - n_{AB}) + (n_B - n_{AB})\log(n_B - n_{AB}) \\ & -(n - n_A)\log(n - n_A) - (n - n_b)\log(n - n_B)\end{aligned}$$

Appendix B

Table A1. Classification of each community under a specific theme.

a. Male.				
Family	**Crime**	**Action**	**War**	**Plot Narration**
female/characters	police/noun	finds/verb	other/adj	movie/noun
is/verb	money/noun	find/verb	men/noun	begins/verb
be/verb	town/noun	house/noun	group/noun	story/noun
has/verb	take/verb	takes/verb	including/verb	ends/verb
family/noun	local/adj	goes/verb	are/verb	show/noun
father/noun	murder/noun	car/noun	using/verb	scene/noun
tells/verb	gang/noun	tries/verb	team/noun	game/noun
time/noun	found/verb	killed/verb	people/noun	final/adj
life/noun	taken/verb	death/noun	order/noun	end/noun
wife/noun	crime/noun	room/noun	use/verb	last/adj
man/noun	arrested/verb	way/noun	ship/noun	making/verb
mother/noun	working/verb	body/noun	city/noun	music/noun
have/verb	case/noun	dead/adj	world/noun	movie/noun
new/adj	prison/noun	kill/verb	several/adj	shows/verb
go/verb	taking/verb	escape/verb	help/noun	band/noun
son/noun	office/noun	leaves/verb	war/noun	play/verb
young/adj	officer/noun	sees/verb	explains/verb	playing/verb
home/noun	evidence/noun	kills/verb	crew/noun	following/verb
friend/noun	place/noun	discovers/verb	small/adj	series/noun
love/noun	stolen/verb	gets/verb	plan/noun	song/noun
b. Female				
Family	**Crime**	**Shopping**	**Suicide**	**Plot Narration**
male/characters	find/verb	store/noun	suicide/noun	other/adj
is/verb	finds/verb	park/noun	commit/verb	movie/noun
be/verb	men/noun	middle/noun	pills/noun	begins/verb
father/noun	killed/verb	homeless/adj	sleeping/verb	including/verb
has/verb	police/noun	mall/noun	overdose/noun	school/noun
tells/verb	group/noun	shopping/noun	committing/verb	are/verb
man/noun	kill/verb	grocery/noun	attempted/verb	people/noun

Table A1. *Cont.*

family/noun	car/noun	cashier/noun	commits/verb	story/noun
time/noun	death/noun	shopping/verb	volunteer/verb	team/noun
wife/noun	tries/verb	bench/noun	shaken/verb	world/noun
life/noun	way/noun	department/noun	contemplating/verb	same/adj
new/adj	escape/verb	aged/adj	contemplates/verb	women/noun
have/verb	dead/adj	convenience/noun		sees/verb
mother/noun	using/verb	amusement/noun		high/adj
take/verb	room/noun	clearing/verb		made/verb
get/verb	kills/verb	liquor/noun		making/verb
go/verb	discovers/verb	hardware/noun		many/adj
own/adj	killing/verb			ends/verb
friend/noun	body/noun			called/verb
takes/verb	causing/verb			show/noun

References

1. Quadflieg, S.; Macrae, C.N. Stereotypes and Stereotyping: What's the Brain Got to Do with It? *Eur. Rev. Soc. Psychol.* **2011**, *22*, 215–273. [CrossRef]
2. Koenig, A.M. Comparing Prescriptive and Descriptive Gender Stereotypes about Children, Adults, and the Elderly. *Front. Psychol.* **2018**, *9*, 1086. [CrossRef]
3. Oakes, P.J.; Haslam, S.A.; Turner, J.C. *Stereotyping and Social Reality*; Blackwell: Oxford, UK; Cambridge, MA, USA, 1994.
4. Ito, T.A.; Urland, G.R. Race and Gender on the Brain: Electrocortical Measures of Attention to the Race and Gender of Multiply Categorizable Individuals. *J. Pers. Soc. Psychol.* **2003**, *85*, 616–626. [CrossRef] [PubMed]
5. Ellemers, N. Gender Stereotypes. *Annu. Rev. Psychol.* **2018**, *69*, 275–298. [CrossRef] [PubMed]
6. Abbott Watkins, T. The Ghost of Salary Past: Why Salary History Inquiries Perpetuate the Gender Pay Gap and Should Be Ousted as a Factor Other than Sex. *Minn. Law Rev.* **2018**, *103*, 1041.
7. Charlesworth, T.E.S.; Yang, V.; Mann, T.C.; Kurdi, B.; Banaji, M.R. Gender Stereotypes in Natural Language: Word Embeddings Show Robust Consistency Across Child and Adult Language Corpora of More Than 65 Million Words. *Psychol. Sci.* **2021**, *32*, 218–240. [CrossRef]
8. Malinowska, A. Waves of Feminism. In *The International Encyclopedia of Gender, Media, and Communication*; Ross, K., Bachmann, I., Cardo, V., Moorti, S., Scarcelli, M., Eds.; Wiley: Hoboken, NJ, USA, 2020; pp. 1–7.
9. Mayol-García, Y.; Gurrentz, B.; Kreider, R.M. Number, Timing, and Duration of Marriages and Divorces. 2016. Available online: https://www.census.gov/library/publications/2021/demo/p70-167.html (accessed on 12 April 2022).
10. Eagly, A.H.; Wood, W. Social Role Theory. In *Handbook of Theories of Social Psychology: Volume 2*; SAGE Publications Ltd.: London, UK, 2012; pp. 458–476.
11. Donnelly, K.; Twenge, J.M. Masculine and Feminine Traits on the Bem Sex-Role Inventory, 1993–2012: A Cross-Temporal Meta-Analysis. *Sex Roles* **2017**, *76*, 556–565. [CrossRef]
12. Eagly, A.H.; Nater, C.; Miller, D.I.; Kaufmann, M.; Sczesny, S. Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of U.S. Public Opinion Polls from 1946 to 2018. *Am. Psychol.* **2020**, *75*, 301–315. [CrossRef]
13. Haines, E.L.; Deaux, K.; Lofaro, N. The Times They Are A-Changing . . . or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychol. Women Q.* **2016**, *40*, 353–363. [CrossRef]
14. Charlesworth, T.E.S.; Banaji, M.R. Patterns of Implicit and Explicit Stereotypes III: Long-Term Change in Gender Stereotypes. *Soc. Psychol. Personal. Sci.* **2022**, *13*, 14–26. [CrossRef]
15. Fiske, S.T.; Linville, P.W. What Does the Schema Concept Buy Us? *Pers. Soc. Psychol. Bull.* **1980**, *6*, 543–557. [CrossRef]
16. Ward, L.M.; Hansbrough, E.; Walker, E. Contributions of Music Video Exposure to Black Adolescents' Gender and Sexual Schemas. *J. Adolesc. Res.* **2005**, *20*, 143–166. [CrossRef]
17. Adoni, H.; Mane, S. Media and the Social Construction of Reality: Toward an Integration of Theory and Research. *Commun. Res.* **1984**, *11*, 323–340. [CrossRef]
18. Durkheim, E. *Sociology and Philosophy (Routledge Revivals)*; Routledge: London, UK, 2009.
19. Stella, M. Text-Mining Forma Mentis Networks Reconstruct Public Perception of the STEM Gender Gap in Social Media. *PeerJ Comput. Sci.* **2020**, *6*, e295. [CrossRef] [PubMed]
20. Bhatia, N.; Bhatia, S. Changes in Gender Stereotypes Over Time: A Computational Analysis. *Psychol. Women Q.* **2021**, *45*, 106–125. [CrossRef]
21. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural Language Processing (Almost) from Scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
22. Siew, C.S.Q.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics. *Complexity* **2019**, *2019*, e2108423. [CrossRef]

23. Xu, H.; Zhang, Z.; Wu, L.; Wang, C.-J. The Cinderella Complex: Word Embeddings Reveal Gender Stereotypes in Movies and Books. *PLoS ONE* **2019**, *14*, e0225385. [CrossRef]
24. Barthes, R.; Duisit, L. An Introduction to the Structural Analysis of Narrative. *New Lit. Hist.* **1975**, *6*, 237–272. [CrossRef]
25. Cuddon, J.A. *The Penguin Dictionary of Literary Terms*; Prentice Hall: Hoboken, NJ, USA, 2005.
26. Sandberg, S. The Importance of Stories Untold: Life-Story, Event-Story and Trope. *Crime Media Cult. Int. J.* **2016**, *12*, 153–171. [CrossRef]
27. Wickham, H. Rvest: Easily Harvest (Scrape) Web Pages. 2021. Available online: https://rvest.tidyverse.org/ (accessed on 12 April 2022).
28. Wikipedia: Consensus. Wikipedia. 2022. Available online: https://en.wikipedia.org/wiki/Wikipedia:CONSENSUS (accessed on 12 April 2022).
29. Benoit, K.; Matsuo, A. Spacyr: Wrapper to the "SpaCy" "NLP" Library. 2020. Available online: https://CRAN.R-project.org/package=spacyr (accessed on 12 April 2022).
30. Honnibal, M.; Montani, I. SpaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing. Available online: https://sentometrics-research.com/publication/72/ (accessed on 12 April 2022).
31. Wais, K. Gender Prediction Methods Based on First Names with GenderizeR. *R J.* **2006**, *8*, 17–37. [CrossRef]
32. Strømgren, C. Genderize.Io. Available online: https://genderize.io/ (accessed on 12 April 2022).
33. Csardi, G.; Nepusz, T. The Igraph Software Package for Complex Network Research. *InterJournal Complex Syst.* **2006**, *1695*, 1–9.
34. Almende, B.V. *Benoit Thieurmel and Titouan Robert VisNetwork: Network Visualization Using "vis.Js" Library*; DataStorm: Paris, France, 2021.
35. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast Unfolding of Communities in Large Networks. *J. Stat. Mech.* **2008**, *2008*, P10008-12. [CrossRef]
36. Bordag, S. A Comparison of Co-Occurrence and Similarity Measures as Simulations of Context. In *Proceedings of the Computational Linguistics and Intelligent Text Processing*; Gelbukh, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 52–63.
37. Sitter, M. Violence and Masculinity in Hollywood War Films during World War II. Master's Thesis, Library and Archives Canada, Lakehead University, Thunder Bay, ON, Canada, 2012.
38. Suicide Death Rate U.S. by Gender 1950–2018. Available online: https://www.statista.com/statistics/187478/death-rate-from-suicide-in-the-us-by-gender-since-1950/ (accessed on 12 April 2022).
39. Levi-Belz, Y.; Gvion, Y.; Apter, A. Editorial: The Psychology of Suicide: From Research Understandings to Intervention and Treatment. *Front. Psychiatry* **2019**, *10*, 214. [CrossRef] [PubMed]
40. Vijayakumar, L. Suicide in Women. *Indian J. Psychiatry* **2015**, *57*, 233. [CrossRef] [PubMed]
41. Zhang, J.; Mckeown, R.E.; Hussey, J.R.; Thompson, S.J.; Woods, J.R. Gender Differences in Risk Factors for Attempted Suicide among Young Adults: Findings from the Third National Health and Nutrition Examination Survey. *Ann. Epidemiol.* **2005**, *15*, 167–174. [CrossRef] [PubMed]
42. Benshoff, H.M.; Griffin, S. *America on Film: Representing Race, Class, Gender, and Sexuality at the Movies*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
43. Winlow, S. Masculinities and Crime. *Crim. Justice Matters* **2004**, *55*, 18–19. [CrossRef]
44. Reed, S.M. Boys to Men: Masculinity, Victimization, and Offending. Master's Thesis, University of Nevada, Reno, NV, USA, 2018; p. 3316. [CrossRef]
45. Spitzberg, B.H.; Cadiz, M. The Media Construction of Stalking Stereotypes. *J. Crim. Justice Pop. Cult.* **2002**, *9*, 128–149.
46. Hall, D.M. The Victims of Stalking. In *The Psychology of Stalking*; Meloy, J.R., Ed.; Academic Press: San Diego, CA, USA, 1998; pp. 113–137.
47. Eschholz, S.; Bufkin, J. Investigating the Efficacy of Measures of Both Sex and Gender for Predicting Victimization and Offending in Film. *Sociol. Forum* **2001**, *16*, 655–676. [CrossRef]
48. Parker, K.; Stepler, R. As U.S. Marriage Rate Hovers at 50%, Education Gap in Marital Status Widens. Pew Research Center. Available online: https://www.pewresearch.org/fact-tank/2017/09/14/as-u-s-marriage-rate-hovers-at-50-education-gap-in-marital-status-widens/ (accessed on 14 September 2017).
49. Empey, L.T. Role Expectations of Young Women Regarding Marriage and a Career. *Marriage Fam. Living* **1958**, *20*, 152. [CrossRef]
50. Badore, A. Gender of a Nation: Propaganda in World War II and the Atomic Age. Available online: https://www.semanticscholar.org/paper/Gender-of-a-Nation%3A-Propaganda-in-World-War-II-and-Badore-Angela/59239a4035c722818f10496d569d87872b517148#paper-header (accessed on 12 April 2022).
51. Morrison, D. Brave: A Feminist Perspective on the Disney Princess Movie. Bachelor's Thesis, California Polytechnic State University, San Luis Obispo, CA, USA, June 2014.
52. Barber, M. Disney's Female Gender Roles: The Change of Modern Culture. Ph.D. Thesis, Indiana State University, Terre Haute, IN, USA, 2016.
53. Fry, R.; Cohn, D. Women, Men and the New Economics of Marriage. Pew Research Center's Social & Demographic Trends Project; Pew Research Center's Social & Demographic Trends Project. Available online: https://www.pewresearch.org/social-trends/2010/01/19/women-men-and-the-new-economics-of-marriage/ (accessed on 19 January 2010).
54. Powers, S.P.; Rothman, D.J.; Rothman, S. Transformation of Gender Roles in Hollywood Movies: 1946–1990. *Polit. Commun.* **1993**, *10*, 259–283. [CrossRef]

55. Horowitz, J.M.; Graf, N.; Livingston, G. Marriage and Cohabitation in the U.S. Pew Research Center's Social & Demographic Trends Project; Pew Research Center's Social & Demographic Trends Project. Available online: https://www.pewresearch.org/social-trends/2019/11/06/marriage-and-cohabitation-in-the-u-s/#:~{}:text=As%20more%20U.S.%20adults%20are,new%20Pew%20Research%20Center%20survey (accessed on 6 November 2019).
56. Oliver, K. The Male Gaze Is More Relevant, and More Dangerous, than Ever. *New Rev. Film Telev. Stud.* **2017**, *15*, 451–455. [CrossRef]
57. Rhode, D.L. Appearance as a Feminist Issue. In *Body Aesthetics*; Irvin, S., Ed.; Oxford University Press: London, UK, 2016; pp. 81–93.
58. Spencer, S.J.; Steele, C.M.; Quinn, D.M. Stereotype Threat and Women's Math Performance. *J. Exp. Soc. Psychol.* **1999**, *35*, 4–28. [CrossRef]
59. Kaye, L.K.; Pennington, C.R. "Girls Can't Play": The Effects of Stereotype Threat on Females' Gaming Performance. *Comput. Hum. Behav.* **2016**, *59*, 202–209. [CrossRef]
60. Bell, A.E.; Spencer, S.J.; Iserman, E.; Logel, C.E.R. Stereotype Threat and Women's Performance in Engineering. *J. Eng. Educ.* **2003**, *92*, 307–312. [CrossRef]
61. Peterson, S.B.; Lach, M.A. Gender Stereotypes in Children's Books: Their Prevalence and Influence on Cognitive and Affective Development. *Gend. Educ.* **1990**, *2*, 185–197. [CrossRef]
62. Bandura, A. Social Learning Theory of Aggression. *J. Commun.* **1978**, *28*, 12–29. [CrossRef]
63. Dill, K.E.; Thill, K.P. Video Game Characters and the Socialization of Gender Roles: Young People's Perceptions Mirror Sexist Media Depictions. *Sex Roles* **2007**, *57*, 851–864. [CrossRef]
64. Saleem, M.; Anderson, C.A. Arabs as Terrorists: Effects of Stereotypes within Violent Contexts on Attitudes, Perceptions, and Affect. *Psychol. Violence* **2013**, *3*, 84–99. [CrossRef]
65. Krys, K.; Capaldi, C.A.; van Tilburg, W.; Lipp, O.V.; Bond, M.H.; Vauclair, C.-M.; Manickam, L.S.S.; Domínguez-Espinosa, A.; Torres, C.; Lun, V.M.-C.; et al. Catching up with Wonderful Women: The Women-Are-Wonderful Effect Is Smaller in More Gender Egalitarian Societies. *Int. J. Psychol.* **2018**, *53*, 21–26. [CrossRef]
66. Saeed, A. Media, Racism and Islamophobia: The Representation of Islam and Muslims in the Media: The Representation of Islam and Muslims in the Media. *Sociol. Compass* **2007**, *1*, 443–462. [CrossRef]
67. Soto-Perez-de-Celis, E. Social Media, Ageism, and Older Adults during the COVID-19 Pandemic. *EClinicalMedicine* **2020**, *29–30*, 100634. [CrossRef] [PubMed]
68. Mocarski, R.; King, R.; Butler, S.; Holt, N.R.; Huit, T.Z.; Hope, D.A.; Meyer, H.M.; Woodruff, N. The Rise of Transgender and Gender Diverse Representation in the Media: Impacts on the Population. *Commun. Cult. Crit.* **2019**, *12*, 416–433. [CrossRef]
69. IMDb. Plots. Available online: https://help.imdb.com/article/contribution/titles/plots/G56STCKTK7ESG7CP# (accessed on 12 April 2022).
70. Navarro, J.G. U.S. & Canada: Movie Releases per Year from 2000 to 2021. Available online: https://www.statista.com/statistics/187122/movie-releases-in-north-america-since-2001/ (accessed on 12 April 2022).
71. Navarro, J.G. Box Office Revenue in the U.S. and Canada from 1980 to 2021. Available online: https://www.statista.com/statistics/187069/north-american-box-office-gross-revenue-since-1980/ (accessed on 12 April 2022).
72. Johns, M.; Schmader, T.; Martens, A. Knowing Is Half the Battle: Teaching Stereotype Threat as a Means of Improving Women's Math Performance. *Psychol. Sci.* **2005**, *16*, 175–179. [CrossRef] [PubMed]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

BDCC Editorial Office
E-mail: actuators@mdpi.com
www.mdpi.com/journal/actuators

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-0365-4346-8