A Novel Deep Neural Network-Based Approach to Measure Scholarly Research Dissemination Using Citations Network

Naif Radi Aljohani 1,*, Ayman Fayoumi 1 and Saeed-Ul Hassan 2

1 Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; afayoumi@kau.edu.sa
2 Department of Computing and Mathematics, Manchester Metropolitan University, Manchester M15 6BH, UK; s.ul-hassan@mmu.ac.uk
* Correspondence: nraljohani@kau.edu.sa

Abstract: We investigated the scientific research dissemination by analyzing the publications and citation data, implying that not all citations are significantly important. Therefore, as alluded to existing state-of-the-art models that employ feature-based techniques to measure the scholarly research dissemination between multiple entities, our model implements the convolutional neural network (CNN) with fastText-based pre-trained embedding vectors, utilizes only the citation context as its input to distinguish between important and non-important citations. Moreover, we speculate using focal-loss and class weight methods to address the inherited class imbalance problems in citation classification datasets. Using a dataset of 10 K annotated citation contexts, we achieved an accuracy of 90.7% along with a 90.6% f1-score, in the case of binary classification. Finally, we present a case study to measure the comprehensiveness of our deployed model on a dataset of 3100 K citations taken from the ACL Anthology Reference Corpus. We employed state-of-the-art graph visualization open-source tool Gephi to analyze the various aspects of citation network graphs, for each respective citation behavior.

Keywords: citations context classification; citation network analysis; deep learning

1. Introduction

Citation analysis has been an active area of research to measure the impact of scientific publications [1–4]. Since the inception of the internet, interactions and communication among scholars around the world have expanded. This presents the concept of knowledge flow, which is the exchange of knowledge among the different scientific communities [5–7]. Researchers from similar research areas tend to be interested in the same articles. Therefore, many techniques have used citation-based networks to improve document recommendation systems [8–10]. Agarwal et al. [11] addressed the problems of high dimensionality input data and real-time consumption by introducing a scalable subspace clustering algorithm (SCuBA) approach. They tested the proposed algorithm on the data corpus from MovieLens (movie ratings by millions of users) and concluded that the approach outperforms subsequent clustering techniques (i.e., fallback models) at 15% precision, and is also faster, scalable, and produces high-quality recommendations.

Similarly, Gori and Pucci [12] used citation-based graphs to construct a research article recommendation system. They proposed a PageRank algorithm that combines the properties of a citation graph and random walk. The algorithm assigns a preference score to a set of publications and linked documents based on their bibliographic references. They experimented with nine datasets and observed that the model ranked the documents in every dataset on a citation graph at the first 20 positions. Küçüktunç et al. [13] designed a search-based recommendation system to find the most recent and relevant documents in response to a user query. They suggested a direction-aware network where each edge has a citation direction, using the dataset from CiteSeer and DBLP and comparing the proposed...
technique with the page rank algorithm. According to the results, PaperRank leads to higher accuracy when the query is generic, whereas a direction-aware method performs better when the query is specific.

Some studies present prototypes that allow the user to visualize the citation network in terms of how other documents use a certain document. Berger et al. [14] presented the “Cite2vec” visualization scheme that permits users to browse documents by using word embedding. In contrast to the above approaches, this technique considers contextual information, as well as the publication venue and author details. Another study, by Ganguly and Pudi [15], presented “Paper2vec”, an approach focused on the neighboring nodes of publications in a citation network. Ebesu and Fang [16] incorporated author information for context-aware recommendations by employing a neural citation network. In this paper, we designed a citation context classification approach using customized focal-loss and class-weight-aware convolutional neural network (CNN) network. This type of technology can be applied for mining social media structured and semi-structured data sources as well [17–21].

We propose that visualizing scholarly research influence can capture multiple perspectives that provide both qualitative and quantitative data. Visualizations are frequently utilized to engage both novice and experienced users. Visualizations can help people and researchers to understand patterns and correlations in datasets and serve as storytelling tools. A well-designed visualization can also assist analysis in varying degrees of detail, ranging from offering a gestalt perspective of a scholarly research study to its dissemination into different disciplines.

The primary contribution of this paper is a broadly accessible, automated approach to visualizing the scientific research dissemination over time. In this paper, we measured the scientific research dissemination by analyzing the Association of Computational Linguistics (ACL) data corpus. We designed a novel focal loss and weight-aware deep neural network for citation context classification. Later, we employed our model to automatically assign the citation function labels to the ACL Anthology Reference Corpus (ARC). Finally, we used Gephi to analyze the various aspects of citation network graphs, for each respective citation behavior. Following are the key contributions of this work.

− Firstly, we prepared a tagged dataset of 9518 citing sentences from the Association for Computational Linguistics (ACL) anthology of research articles. We provided a seven-class annotation method for citation classification by combining the citation schemes proposed by Jurgens et al. [22] and Teufel et al. [23]. For citation classification, we presented a seven-class annotation scheme: use, extend, compare & contrast, motivation, background, future, and none. Later, these seven classes were divided into two categories: important (uses and extends) and non-important citations (compare or contrast, motivation, background, future, and none).
− Secondly, to tackle the fundamental problem of class imbalance in citation text classification tasks, we presented a customized focal-loss and class-weight-aware convolutional neural network (CNN) network. Moreover, we argue that adopting the customized focal-loss method increases the overall cross-entropy function by a factor of $(1 - p_t)^\gamma$ and that the proposed CNN outperforms current state-of-the-art techniques.
− Lastly, we deployed our classification model on the citation network of ACL Anthology Reference Corpus (ARC) to automatically assign citation functions (use, extend, compare or contrast, motivation, background, future) against each citation. The ARC corpus comprises 318,351 citations instance that makes a network between publications in the ACL Anthology and citations from outside the ACL Anthology with canonicalized IDs. We used the Gephi to analyze the various aspects of citation network graphs, for each respective citation behavior.

2. Literature Review
A citation network is a graph, its nodes being articles and its edges citation links between them [24,25]. In recent years, a number of studies have been conducted to leverage
deep neural networks and word embedding to extract entities from web and social network datasets [26–30].

Liu et al. [31] covered the problem of detecting interlinked author communities within the citation networks and the topics covered by the documents in a collection of data was addressed. They employed the topic-link LDA model (for the formation of network links, topic similarity, and community similarity are measured) to uncover networks of the authors and topics using citation data from CiteSeer. The results demonstrate that the proposed approach outperforms the content-based approach, at a 7% F1 measure. Similarly, Eto [32] proposed a co-citation search method that is a combination of graph-based algorithm and co-citation context, comparing it to the traditional co-citation search method without context. For experimentation, they used biomedicine and computer linguistics datasets with a total of 172,734 and 10,921 documents, respectively, and stated that the model with citation context performs better than it does without.

Chang and Blei [33] predicted the links between documents and their summarizations using Relational Topic Modelling (RTM: a data model composed of document words and the links between them). The authors worked with datasets from three databases: WebKB (contains web pages of computer science departments at several universities); CORA (Coriolis Ocean Database for Reanalysis); and PNAS (Proceedings of the National Academy of Sciences). Their results showed that RTM models predicted the links between documents with improved precision of 5%, compared to unigram and LDA models.

In another study, Yang et al. [34] combined four citation network methods (direct citation, coupling, indirect citation, and co-citation) for a network of multiple types of citation. The results verify that the combination reveals complex information about citations and locates a greater number of influential citations than direct citation-based networks. A comparison of citation-context networks has been performed by Bornmann, Haunschild, and Hug [35]. They retrieved the citation context of articles from the WoS repositories and, using VOSviewer, generated various co-occurrence networks based on articles’ keywords, abstracts, and titles. The results show that networks based on keywords from titles and abstracts show the broad range of the research areas covered by the citing article, while networks based on context reveal the cognitive impact of the cited article on the citing work.

Citation recommendation systems can be divided into two categories. The first group is global systems, which take the entire scholarly article as their input and recommend the citations for the article. Meng et al. [36] proposed a citation recommendation approach that incorporates citation context, the article’s content, and its authorship. They conducted experiments on 12,762 articles published before 2012 from the ACL Anthology Network. The results demonstrated that the suggested recommendation system achieves about 16.2% more MAP than a system that is simply based on query text. Kong et al. [37] designed a recommendation system named “VOPRec”; they considered both textual information and network topologies simultaneously. To find similar research topics, textual information is converted into word embedding. Using physics journals as a dataset, the experiments showed that the proposed system achieved better precision, recall, and F1 than the state-of-the-art models (i.e., Doc2vec and Struc2vec).

Some studies have presented novel techniques for efficient global, as well as context-aware, recommendation systems. Zhu et al. [38] designed the citation recommendation system “ActiveCite”, which uses three approaches: collaborative filtering, content-based analysis, and citation analysis. The proposed system automatically employs the citation context as a query in local citation recommendations and mines articles’ subjects as queries in context-aware recommendation tasks to recommend related reference sources.

Scientific impact measurements are traditionally based on the total number of citations. Fang [39] investigated the impact of a researcher by presenting an automated system to count both direct (to include the author’s surname as part of the text and year of publication in parentheses) and indirect citations (to include both the author’s surname and year of publication in parentheses) of a publication. According to him, to assess a researcher’s scientific impact one needs to sum the citations of all articles. Using medical articles from
1991 to 2001 from the WoS database, Patsopoulos et al. [40] measured the annual citation count of each publication and concluded that knowledge is rarely consumed in the same year in which it is produced: a publication achieves a representative citation count two years after its publication. Similar results were reported by Lu, Ding, and Zhang [41], who investigated the influence of a highly cited article on those citing articles published between 2006 and 2014 and indexed in the WoS database. They proposed using the citation context with specific features (citation location, citation mention, and citation topic) and observed the change in articles’ impact over a nine-year interval after publication. They concluded that the consumption of H-index articles peaked in the first two years, until 2007, and then started declining, with fluctuations.

Zhuge [42] claimed that the concepts or ideas introduced in publications lead to new ideas that will, in turn, be published in articles on similar topics soon. Therefore, citations between publications imply a knowledge flow from the author of the article being cited to the author of the article that cites it. Another approach, by Zhang et al. [43], deployed the citation influence model to reveal the production and consumption of knowledge in the field of physics across various countries. They analyzed the studies published in the American Physical Society (APS) over a 50-year time interval (1960–2009). The results concluded that, in the early years, the United States was the top producer of knowledge in the field of physics, whereas in the early 1990s the United Kingdom and northern Europe were the dominant producers. Throughout this time, China was the largest consumer of knowledge worldwide. Many studies have presented novel measures to calculate the spread of scientific knowledge, such as the Rowlands Diffusion Index [44] and the Frandsen Diffusion Index [45]. Hassan and Haddawy [46] proposed a novel metric to measure the citation impact of countries by analyzing Elsevier’s Scopus citation database in the subject area of energy for the years 1996 to 2009. The proposed metric is entitled the International Scholarly Impact of Scientific Research (ISISR), and it calculates the number of citations of an article in a particular domain by researchers from both beyond and within the source country. The results indicate that China produced more publications in the field of energy, but that they were not referenced by the international research community, whereas the United States produced fewer publications that are referenced internationally.

Researchers have used various publication sets to measure the diffusion of knowledge between several units, such as among subject categories, journals, institutions, scientists, and countries. To investigate the international visibility of journals, Zhou and Leydesdorff [47] adopted a journal-citation analysis approach. Using two databases, Science Citation Index (SCI) and China Scientific and Technical Papers and Citations Database (CSTPCD), they compared the citation count of Chinese articles to that of international publications. The results showed that Chinese articles have a lower citation count since Western articles are written in English, which is understood by every author. Although Chinese journals are accessible to international authors, the Chinese language prevents them from understanding the content.

There have been efforts to calculate knowledge flows using semantic analysis. Hassan and Haddawy [48] semantically analyzed knowledge flow by proposing a topic model with a distance matrix, using the Scopus database in the subject area of energy during the years 2004 to 2009. They compared Japanese and Chinese articles that cite literature published in the United States and noted that the former focused on domains such as energy conversion and the efficient use of photovoltaic and superconductors, whereas the latter focused on power grids, power systems, and solar cell production. In the same field, Qasim et al. [49] used the publication and citation data indexed in Scopus to present an approach to studying the creation and the consumption of scientific knowledge relevant to sustainable and renewable energy produced by the United States from 1996 to 2009 across several regions. The results demonstrate that by consuming the knowledge produced by the United States, the Japanese scientific community focuses on topics strongly related to the production of hybrid and electronic vehicles and fuel cells, whereas the Chinese community focuses on renewable energy and biomass. More recently, Hassan et al. [50]
analyzed publication and citation data to measure the knowledge flow between institutions and countries. They investigated the references from PLOS ONE journal in the field of computer and information science.

3. Methodology

In this section, we present the implementation details of our deep-learning citation context classification approach, along with the used pre-trained embedding vectors such as fastText to construct the CNN [51,52] model. To deal with the data imbalance, we proposed employing two approaches to maximize the model’s accuracy: class-weight and focal-loss functions. Furthermore, we deployed the citation classification model on the ARC citation network to automatically label the citation functions against each citation between ACL and the external source papers (see Figure 1).

![Model architecture for citation context classification with added loss-calculation approaches.](image)

3.1. Dataset

We prepared a tagged dataset of 9518 citing sentences from the Association for Computational Linguistics (ACL) anthology of research articles. We provided a seven-class annotation method for citation classification by combining the citation schemes proposed by Jurgens et al. [22] and Teufel et al. [23]. For citation classification, we presented a seven-class annotation scheme: use, extend, compare or contrast, motivation, background, future, and none. Later, these seven classes are divided into two categories: important and non-important citations.

3.2. Deep-Learning-Based Citation Classification Approach

In this section, we present the details of our deep neural network for citation context classification. We proposed a CNN [53,54] model by employing fastText pre-trained embedding vectors. To handle the data imbalance, we employed using two different state-of-the-art techniques such as class-weight and customized focal loss function to improve the model accuracy.

FastText (https://github.com/facebookresearch/fastText) is a widely used library, designed by Facebook. It has permitted researchers and text classifiers to efficiently learn the text representations for text classification. Word embedding is an important part of acquiring language semantics in natural language processing. Many researchers are investigating the usage of Word2Vec [55] vectors to put the comparable words to neighboring points on a plane. Although this technique is capable of learning text patterns, it is incapable of learning unusual word representations because of its entire reliance on a preexisting textual dictionary collection. We employed character-level word embedding, such as fastText, to solve this problem.
FastText achieves astonishing results in learning word representations and sentence categorization by exploiting character-level information, especially in the case of rare terms. The n-gram vectors and the word itself are used to describe each word. Additionally, FastText is an aggregation model where word vectors are regarded as the aggregate of their character-level embedding vectors, rather than simply the word itself, and may thus be used to a variety of text problems. FastText works by using character-level embedding vectors of each n-gram character, instead of explicitly understanding a feature vector for a word (as Word2vec does). Every word in the dictionary is characterized as a sack of n-grams, and the ultimate word vector is the combination of these n-grams.

Moreover, fastText offers text embedding vectors that have been pre-trained using 157 languages on substantially bigger training datasets using the fastText approach. Their use can save training time while also allowing for semantically rich embedding vectors, resulting in better classification results. Instead of using basic Word2Vec, we used the million-word vector file given by fastText for citation context categorization, which was trained on the Wikipedia 2017, UMBC web base corpus, and statmt.org news datasets.

### 3.2.1. CNN Architecture and Citation Context Representation

Furthermore, to categorize the context of citations, we designed a CNN network. Figure 1 depicts the model’s architecture. Let $S$ be the sentence to be categorized, and $x_i$ denotes a word from $S$ with a length of $n$, whereas $k$ is the $k$-dimensional vector for sentence $S$, and $x_i$ to be a $k$-word.

$$S = x_{1+n} = |x_1 + x_2 + x_3 + x_4 + \ldots + x_n|$$ (1)

where $||$ represents the concatenation of all $S$ words. In general, $x_{i:i+j}$ denotes the concatenation of words starting with $i$ and ending with $j$.

$$x_{i:i+j} = |x_i + x_{i+1} + x_{i+2} \ldots x_{i+j}|$$ (2)

The $w$ is the filter projected to a window of $j$ words to obtain a single list of features in a convolution layer. A window of words $x_{i:i+j-1}$ is used to construct feature $F$, using Equation (3) as follows:

$$F = f(w \cdot x_{i:i+j-1} + b)$$ (3)

A max-pooling method is then used to find the maximum number of features from a feature map $c$ for a specific filter $F$, using Equation (4) as follows:

$$\hat{c}_i = \max(c_i)$$ (4)

The goal of max pooling is to find the most visible and significant features having the maximum fmap score. The pooling function has the capacity to handle the forward sentences varying lengths by default. We employed two 1D convolution layers with 64 filters of varied sizes, to extract several features. We used the Relu function to incorporate non-linearity after the first 1D, and then again, a max-pooling layer is added (see Equations (5) and (6)). After that, a global max pooling is called out to the second layer of 1D convolution layer having 64 filters (see Equations (7) and (8)).

$$y_i^{(1)} = ReLU(c_i)$$ (5)

$$y_i^{(1)} = \max(y_i^{(1)})$$ (6)

$$y_i^{(2)} = ReLU(W^{(2)}y_i^{(1)} + b^{(2)})$$ (7)

$$y_i^{(2)} = \max(y_i^{(2)})$$ (8)
Moreover, feature sets from the concluding layers are added to an FC layer of neurons, where the sigmoid activation function is used to calculate the probability value for each class (see Equations (9) and (10)).

\[ y^{(3)} = W^{(2)} y^{(1)} + b^{(2)} \]  

\[ p(i) = \frac{\exp(y_k^{(2)})}{\exp(y_k^{(2)}) + 1} \]  

To mitigate model overfitting, we used regularization to reduce correlated learning among the neurons by placing a penalty factor on the loss computation function. Dropout is a well-known regularization technique; thus, we used it on the outermost layer to limit the weight vector’s L2 norm. As a result, we used the dropout mechanism for the unit \( y \) in forwarding pass (see Equation (11)) instead of conventional weighting (see Equation (12)).

\[ y = w \cdot z + b \]  

\[ y = w \cdot (z \times \delta) + b \]  

where \( \times \) stands for attribute-wise multiplication of \( z \) with the dropout rate \( \delta \), \( \delta \sim \text{Bernoulli}(p) \), and \( \delta = 1 \) for probability \( p \) and 0 in all other situations. A gradient descent strategy is employed during back propagation.

3.2.2. Techniques to Handle the Class Imbalance Problem

In this section, we tackle the problem of imbalanced classes by introducing two sophisticated classifier-level data balancing strategies that alter the model training procedure while retaining the training of the original datasets. We used the weight-balancing function and focal-loss functions at the classifier level instead of rule-based data-level balancing techniques such as random undersampling and oversampling. The mechanisms used are discussed in detail in the following paragraphs.

**Weight balancing** is a method for balancing datasets by altering the weightage measure for each class sample during the loss function computation. In most cases, every sample and every class are given equal and similar weightage; however, in the case of relevant classes, we might want to give certain of them high weightage. Similarly, in the case of the important and non-important binary classes of citation functions, instead of wasting time and resources looking for more data samples for the minority and relevant class that only makes up 10% of our dataset, we deployed weight balancing to both our classes (binary classes), giving the “important” class a higher score to reflect its greater significance.

**Focal loss** is a procedure in which hard-classified examples are given more weight than well-classified instances. We usually have some data samples in our dataset that are simple to classify. Although, these data samples were separated with ~99% efficacy in training, unlike the more challenging and complicated data instances, yet the outcomes were unsatisfactory. The main problem here is that the readily classified training samples provided just as much to the total cross-entropy loss value as the more difficult data points that improved our overall accuracy when correctly identified, and hence to which more emphasis should be given.

Our multi-class data comprises seven classes, the most common of which are background and none, resulting in well-classified data instances. To deal with the data imbalance, we used a focal-loss function, which reduces the weight of well-classified samples while increasing the weight of minority-class samples, which are more difficult to classify. The generic cross-entropy function is shown in Equation (13) as follows:

\[ \text{Cross Entropy} = -\log(p_i) \]  

\[ \text{Focal Loss} = -(1 - p_i)^\gamma \log(p_i) \]
where \( p_t = \begin{cases} p_y & \text{if } p_y = 1 \\ 1 - p & \text{otherwise} \end{cases} \).

Equation (14) depicts the focal-loss function, which boosts the general cross-entropy function by a factor of \((1 - p_t)\gamma\).

### 3.3. Evaluation of the CNN-Based Citation Classification Model

In this section, we present the findings for both binary and multi-class classification. Table 1 presents the precision, recall, f1-score, and accuracy results for citation context classification using both binary and multiclass schemes. Firstly, we used the entire dataset of 9518 cases for binary classification, with 1007 goes to the “important” class and the remaining 8511 to the “non-important” class. To address the problem of class imbalance in our dataset, we applied weight balancing to both binary classes, with the “important” class receiving more weight. Over the course of 60 epochs, our model improved from 53% accuracy to 98% training accuracy. The CNN model performed well in validation testing, with an accuracy of roughly 90%. Our model achieved a 0.906 accuracy score and a 0.906 f-measure. The precision-recall curve for binary classification is shown in Figure 2a, illustrating that both curves represent a similar stable trend.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Citation Functions Scheme</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation Context Classification Using CNN with FastText</td>
<td>Binary-Class</td>
<td>0.906</td>
<td>0.904</td>
<td>0.905</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>Multi-Class</td>
<td>0.787</td>
<td>0.640</td>
<td>0.70</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Secondly, we divided our 9518 occurrences into seven categories for the multi-classification (use, extend, background, motivation, compare or contrast, future, none). To address the data imbalance problem, we utilized a customized loss function in training to suppress the well-classified samples and emphasized more on the minority-class samples, which are more difficult to classify. Equation (14) presents the loss function formula, which increases the general cross-entropy function by a factor of \((1 - p_t)\gamma\).

Our multi-class model began with 0.33 model accuracy and gradually improved to 0.98 over the span of 60 epochs, whereas the multi-class CNN model attained roughly 72% accuracy in validation. The PR curves are shown in Figure 2b, in which both of the precision and recall curves show a consistent increase in precision and recall values.
4. Case Study: Citation Network Analysis Using ACL Data Corpus

As is discussed in this section, we investigated the behavior of the ACL research community, in the light of the scientific community’s well-known needs and growing interest in the expansion of social network analysis platforms to exploit traditional, citation-based scientific research assessments. In this paper, we looked at the main communities with respect to the multiple citations behavior they exhibit using the ACL ARC citation network.

4.1. Dataset and Method

We used the ACL ARC citation network data comprised of 318,351 citations instances (edges) and 86,825 papers (nodes) that make a citation network between publications in the ACL Anthology and outside the ACL Anthology with canonicalized IDs [22]. ACL dataset is freely available online along with their full-text and citation networks. Therefore, we used this dataset. We further deployed our trained CNN-based citation classification model to automatically classify these citations links into binary class (important and non-important) and multi-class (use, extend, compare or contrast, motivation, background, future) citation functions.

We used the state-of-the-art graph visualization open-source software Gephi [56] to analyze many elements of the citation graph created by the ACL ARC citations dataset between publications from ACL and outside the ACL. Furthermore, we utilized the multiple properties of the graphs, Table 2 presents the abbreviations and the details of the used algorithms:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average degree</td>
<td>AD</td>
<td>The average degree of a graph in a citation network is defined as the average number of citations or edges a node has with other nodes [57].</td>
</tr>
<tr>
<td>The weighted average degree</td>
<td>WAD</td>
<td>It is determined from the average number of connections that a node in a network has with another node, where weight is generally specified as the total number of edges for a given node [57].</td>
</tr>
<tr>
<td>Graph density</td>
<td>GD</td>
<td>A graph is said to be complete if every node is connected to every other node by an edge. This property is taken into account by graph density, which counts the number of edges in a graph to determine how close a graph is to completion.</td>
</tr>
<tr>
<td>Modularity</td>
<td>Mod.</td>
<td>In a graph, modularity is a measure of the strength of nodes that develop modules or clusters in the network. A highly connected group of nodes gives a high modularity value; yet, these groups of nodes may have sparsely interconnected connections [58].</td>
</tr>
<tr>
<td>Connected component</td>
<td>CComp.</td>
<td>A connected component (CComp) is a sub-graph of an undirected graph with a path connecting every other node in that sub-graph [59].</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>CCoef.</td>
<td>The clustering coefficient is a measure of how well all of the nodes in a network cluster together [60].</td>
</tr>
<tr>
<td>Average path length</td>
<td>APL</td>
<td>The longest path between two nodes in a network is referred to as path length. It is worth noting that the Average path length gives the perception of information diffusion in a network [61].</td>
</tr>
<tr>
<td>Eigenvector centrality</td>
<td>EC</td>
<td>According to Bonacich [62], Eigenvector Centrality is the measure of influence that a particular node holds in a network based on its number of connections.</td>
</tr>
</tbody>
</table>

In addition to the quantitative indicators mentioned in Table 2, we highlighted the edges between the nodes (papers) to illustrate the interaction among ACL ARC papers with respect to their citation behaviors, both for binary and multi-class citation functions. Each edge is highlighted by a different color to represent the different citation behaviors.
Overall, we examined our citation network in terms of three aspects: (a) we offered a variety of quantitative metrics to visualize the overall trend of ACL ARC data; (b) we displayed networks of selected papers to demonstrate the community interactions; (c) we illustrated the citation edges on the basis of their citation function using the top 1% nodes with respect to their node degree.

Table 3 presents the network level properties of the citations network using complete ACL citations data and the top 1% of the total nodes. ACL ARC data forms a dense network (nodes: 86,825, edges: 318,351), with an average degree greater than 15 and a modularity value of 0.609. Taking top 1% of the total nodes with respect to node degree, the network looks quite sparse with the average degree greater than 10 and the modularity value of 0.60. Both of these networks possess a quite less value of clustering coefficient.

<table>
<thead>
<tr>
<th>Layout Settings</th>
<th>Type</th>
<th>Nodes</th>
<th>Edges</th>
<th>AD</th>
<th>M</th>
<th>CC</th>
<th>C Co</th>
<th>APL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ForceAtlas2 Noverlap</td>
<td>Community detection</td>
<td>2757</td>
<td>22,049</td>
<td>15.99</td>
<td>0.60</td>
<td>26</td>
<td>0.125</td>
<td>3.53</td>
</tr>
<tr>
<td>OpenOrd</td>
<td>Edge-based analysis</td>
<td>919</td>
<td>4806</td>
<td>10.45</td>
<td>0.60</td>
<td>19</td>
<td>0.14</td>
<td>3.50</td>
</tr>
</tbody>
</table>

4.2. Discussion

In Figure 3, we visualize the ACL ARC citations network that forms eight communities, having more than 5% of total nodes and edges. For better visualization, nodes of total degree <29 (2.5% of the total nodes) are filtered from the citations network. We used Gephi with ForceAtlas2 and Noverlap layout settings for the network visualization. The community colors in the network are assigned automatically using a community-detection algorithm (modularity), and the size of nodes is determined by the betweenness centrality measure, with min size = 10 and max size = 50. The network shows the strong interconnectivity between clusters 18, 8, 4, and cluster 16. Cluster 20 is the largest, covering more than 16% nodes of the total nodes. We also observed that among the 16% nodes (papers) covered by cluster 20, only 46% edges (citations) (including 44% uses and 2% extends) links belong to the important citations. Moreover, in cluster 20, nearly 54% of the total edges and nodes belong to the non-important class, having 20% background, 22% compare or contrast, 7% motivation, and 5% future classes citations.

Figure 3. Visualization of the ACL ARC citation network. The citation network illustrates a color-based community structure, nodes having the same colors belong to the same community.

Figure 4 presents the nodes-wise representation of the ACL ARC citations network. Of the total nodes, only 24% of papers belong to the ACL whereas the remaining 75%
of papers are external papers. The blue color indicates the ACL paper and the red color represents the external papers. It is important to note that in the ACL citations network, the majority of the citations come from outside the ACL.

![Figure 4](image.png)

**Figure 4.** Visualization of the ACL ARC citations network to visualize the ACL papers and papers outside the ACL.

Furthermore, among the total nodes, the top 1% nodes with respect to node degree are taken to visualize the citation network, based on their citation functions such as important (uses, extends) and non-important (background, compare or contrast, motivation, future). Figure 5a illustrates the binary class citations edges between the papers. Of the total citations, only 37.84% of citations are important, and the remaining 62.16% are non-important citations. Similarly, Figure 5b presents the citations edges based on the multi-class citations functions. We find “uses” and “background” classes at the top of the list, with 34.64% and 32.8% citations received, respectively. The motivation, future, and compare or contrast classes have more than 5% citations instances, while the “extends” class received less than 5% citations.

![Figure 5](image.png)

**Figure 5.** Visualization of the ACL ARC citations network to only visualize the citation classes: (a) presents the binary class citations edges and (b) illustrates the multi-class citation edges between the nodes. The top 1% of total nodes with respect to the node degree are taken to visualize the citations network. We used the OpenOrd layout in Gephi to see these citations edges.
5. Concluding Remarks

Analyzing the scientific research influence is a crucial task for researchers and funding organizations. Common methods such as citation counts can overlook a lot of richness and multidimensionality of scholarly research. The scholarly literature is connected by citations and footnotes to form a large citation network. Over the years, this well-preserved structure connects scientific publications, authors, ideas, and multiple disciplines through its billions of citation links. These citations networks can illustrate where ideas have come from and where they may be headed.

In this article, we proposed a novel deep neural network-based technique to measure scientific research dissemination by analyzing the ACL ARC citations network. Our designed CNN-based classification technique, classifying citation contexts into binary (important and non-important) and multi-classes such as background, compare & contrast, extends, uses, motivation and future. Our results indicated that the majority of the cited and citing papers in ACL ARC data belongs to external sources, outside the ACL community. Moreover, we also highlighted that more than 60% of the total citations belong to the uses and background classes.

In conclusion, the deep neural network-based approach used in this study facilitates the assessment of scientific research dissemination. The method utilized is essential to the systems that recognize and analyze the new research themes and measure the impact of scientific publications in the increasingly scholarly big data. In future, using the employed methods, we intend to construct bibliometric enhanced information systems for improved retrieval of bibliographic databases.

Author Contributions: All authors have contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant No. RG-14-611-40. The authors, therefore, gratefully acknowledge DSR’s technical and financial support.

Data Availability Statement: Not applicable.

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

References

41. Lu, C.; Ding, Y.; Zhang, C. Understanding the impact change of a highly cited article: A content-based citation analysis. *Scientometrics* 2017, 112, 927–945. [CrossRef]


46. Hassan, S.U.; Haddawy, P. Measuring international knowledge flows and scholarly impact of scientific research. *Scientometrics* 2013, 94, 163–179. [CrossRef]

47. Zhou, P.; Leydesdorff, L. A comparison between the China Scientific and Technical Papers and Citations Database and the Science Citation Index in terms of journal hierarchies and interjournal citation relations. *J. Am. Soc. Inf. Sci. Technol.* 2007, 58, 223–236. [CrossRef]

48. Hassan, S.U.; Haddawy, P. Analyzing knowledge flows of scientific literature through semantic links: A case study in the field of energy. *Scientometrics* 2015, 103, 33–46. [CrossRef]


