International Journal of *Geo-Information*

MDPI

Article

Modelling and Analyzing the Semantic Evolution of Social Media User Behaviors during Disaster Events: A Case Study of COVID-19

Xuehua Han [1,2] and Juanle Wang [1,3,4,*]

[1] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; hanxh@lreis.ac.cn
[2] Beijing Municipal Institute of City Planning & Design, Beijing 100045, China
[3] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[4] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
* Correspondence: wangjl@igsnrr.ac.cn; Tel.: +86-010-6488-8016

Abstract: Public behavior in cyberspace is extremely sensitive to emergency disaster events. Using appropriate methodologies to capture the semantic evolution of social media users' behaviors and discover how it varies across geographic space and time still presents a significant challenge. This study proposes a novel framework based on complex network, topic model, and GIS to describe the topic change of social media users' behaviors during disaster events. The framework employs topic modeling to extract topics from social media texts, builds a user semantic evolution model based on a complex network to describe topic dynamics, and analyzes the spatio-temporal characteristics of public semantics evolution. The proposed framework has demonstrated its effectiveness in analyzing the semantic spatio-temporal evolution of Chinese Weibo user behavior during COVID-19. The semantic change in response to COVID-19 was characterized by obvious expansion, frequent change, and gradual stabilization over time. In this case, there were obvious geographical differences in users' semantic changes, which were mainly concentrated in the capital and economically developed areas. The semantics of users finally focused on specific topics related to positivity, epidemic prevention, and factual comments. Our work provides new insight into the behavioral response to disasters and provides the basis for data-driven public sector decisions. In emergency situations, this model could improve situational assessment, assist decision makers to better comprehend public opinion, and support analysts in allocating resources of disaster relief appropriately.

Keywords: semantic evolution; spatio-temporal; social media; user behaviors; COVID-19

Citation: Han, X.; Wang, J. Modelling and Analyzing the Semantic Evolution of Social Media User Behaviors during Disaster Events: A Case Study of COVID-19. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 373. https://doi.org/10.3390/ijgi11070373

Academic Editors: Wolfgang Kainz and Wei Huang

Received: 22 April 2022
Accepted: 4 July 2022
Published: 5 July 2022

1. Introduction

The proliferation of location-based services has produced an explosive growth of massive social media data with geographic location information. Social media users are not only receivers but also publishers and disseminators of information. These burgeoning, low-cost, and extensively used 'human sensor' technologies have provided an unprecedented opportunity to understand human behavior and how it varies across space and time [? ? ?]. Human behavior is defined as 'something done, felt, or thought in response to a situation or event, which includes not only direct actions, but elements related to cognitive and perceptive' [?]. Extracting semantics from social media text is very useful for inferring human behavior, potential courses of action, and public opinion [?]. In recent years, more and more researches focus on leveraging the semantic information in social media data to enhance responses to crisis events (e.g., natural disasters, terrorist attacks, and public health events). Researchers have commonly used natural language processing methods to mine social media text and obtain people's opinions [? ? ? ?], analyze public sentiments [?

? ?], and investigate human behavioral patterns [? ? ?]. However, there exists a big challenge about the semantic evolution of social media user behavior and how it varies across geographic space and time due to a lack of appropriate methodologies.

In emergency situations, social media users' behavior is characterized by irrationality, strong infectivity, and conformity. The emergence of massive information during disaster events makes people impressionable and causes their opinions to change as the disaster event develops. Existing research analyzing semantic evolution concentrated on the text mining and information dissemination field, ignoring the geographical and temporal characteristics of users' behavior and lacking fine-grained analysis of topic semantics evolution, specifically for disaster events. There is a critical need to extract the thematic and temporal–spatial patterns of public behavior during disaster events. Taking COVID-19 as a case study, this paper aims to describe the semantic and spatio-temporal evolution of behavioral responses by Weibo users in China. By combining complex networks, topic models, and GIS, this article (1) utilizes a topic extraction and classification model proposed by our previous works [? ?] to extract topics from Weibo texts during the COVID-19 outbreak; (2) builds a user semantic evolution model to describe and analyze the public semantics change; (3) visualizes and analyses the spatio-temporal evolution characteristics of public semantics across the COVID-19 crisis.

2. Related Work

Social media data have been increasingly used to assist with disaster management from a human-centric perspective [? ? ? ?]. Scheele [?] proposed a geographic context-aware text mining method, integrating spatial–temporal text features of social media to classify Hurricane Sandy relevant Twitter posts. Chen [?] used two quantitative metrics—the fraction of event-related tweets (FET) and the net positive sentiment (NPS) to examine the intensity and direction dimensions of public opinion on society reopening amid COVID-19 for decision making in emergency events. Gruebner [?] analyzed the temporal and spatial characteristics of public negative emotions before, during, and after superstorm Sandy from Twitter data. Gongora-Svartzman [?] introduced a method combining text processing techniques and graph network analysis to measure social cohesion through social media networks. Shi [?] constructed a text emotion extraction method based on a dictionary of emotional ontology to analyze the developmental course of online public opinion on Weibo during the COVID-19 epidemic in China. Although an increasing number of studies have focused on semantic extraction from social media to analyze public opinion, emotions, and behaviors during disaster, there is less research on the semantic change in social media over time–space during a disaster event.

2.1. Semantic Extraction and Evolution

Topic models (e.g., latent Dirichlet allocation [LDA], latent semantic analysis) and machine learning (e.g., support vector machine [SVM] and random forest [RF]) have been successfully employed to extract topics from social media, analyze topics varying across space and time, and describe public behaviors. However, studies on the semantic evolution of users' topics are insufficient. Resch [?] applied LDA to extract semantic information from Twitter and performed spatio-temporal analysis to estimate the footmark and damage made by the Napa (CA, USA) earthquake in 2014. Using LDA and sentiment analysis, Dahal [?] analyzed tweets relating to climate change to compare the characteristics of climate change discussions between different countries and time period. Wang [?] utilized LDA and SVM to extract topics from Weibo texts on the 2012 Beijing Rainstorm and investigated the distribution pattern of different topics. Zong [?] analyzed content and temporal characteristics of tweet responses to the explosions of a chemical plant in Tianjin, China in 2015. Burnap [?] forecasted the size and survival of information flow during the Woolwich terrorist event in London, in 2013, by analyzing users' sentiments and the frequency of information posted on Twitter. Ye [?] adopted LDA to classify Weibo posts relevant 'dengue' into five classifications and analyzed the relationship between the online discussion and the outbreak of the disease on Weibo.

Most existing research on topic evolution is in the field of text mining. A probability model, a clustering algorithm, and topic similarity metrics were combined to construct a topic evolution model [?]. Blei and Lafferty [?] put forward a dynamic topic model focusing on the topic composition change, that is, word distributions. Ahmed and Xing [?] introduced an infinite dynamic topic model, which could accommodate the evolution of the number, distribution, and popularity of topics in the latent structure. Han [?] developed topic model supported by adaptive and time-varying metadata and applied it to a large Weibo dataset. He [?] proposed an 'attention shift network' framework to analyze the attention dynamics of crowds related to disasters by tracing Twitter hashtags. However, existing research ignores the spatio-temporal characteristics of social media data and lacks a fine-grained analysis of topic semantic evolution.

2.2. Complex Networks and Social Media

Complex networks are networks with some or all of the following characteristics: self-organization, self-similarity, being scale-free, being an attractor network, or being a small world network [? ?]. The topological structure of a complex network can be represented by a graph, which abstracts the real individuals as nodes and the relationships between individuals as the connected edges between the nodes. A single complex network can be formally defined as G = (N, E), consisting of the nodes set N and the edges set E. An undirected network consists of a set of objects (called vertices or nodes) that are connected together, where all the edges are made bidirectional [?]. In contrast, a network where the edges point in a particular direction is called a directed network.

From the complex network perspective, social media is an example of a complex social network used to express and disseminate users' feelings, attitudes, and opinions [?]. A complex network perspective can achieve a deeper understanding of social media and excavate more potential information [?]. Existing studies are in the information dissemination field, where scholars have established models of information dissemination for rumors, public opinion, and other subjects [? ?]. Some have built social network public opinion information dissemination models, investigated the factors influencing social public opinion communication [? ?], and detected the communities of a social network [? ?]. These studies mainly focused on the spread of public opinion in social networks and understanding the mechanisms and factors affecting information dissemination. However, there is little work for systematic analysis of the social media users' topics evolution combined complex networks and GIS.

3. Data and Methods

3.1. Technical Framework

Figure ?? illustrates the framework for extracting the thematic evolution and temporal–spatial pattern of social media behavior. The framework mainly consists of three parts as follows: (1) topic extraction. We collected Weibo texts for a certain period centered on an event, after text filtering, LDA model, and random forest algorithm were each implemented to hierarchically derive topics and sub-topics [?]. (2) Semantic change model. Based on complex network, the users' sub-topics were transferred to semantic change datasets, identifying topic pairs as nodes and edges. In addition, the semantic change dataset was partied by a proper time windows to build the semantic change networks per time window. (3) Visualization and analysis. We achieved visualization of all the semantic change networks, analyzed thematic evolution and network structure change, and performed spatiotemporal analysis of user semantic change.

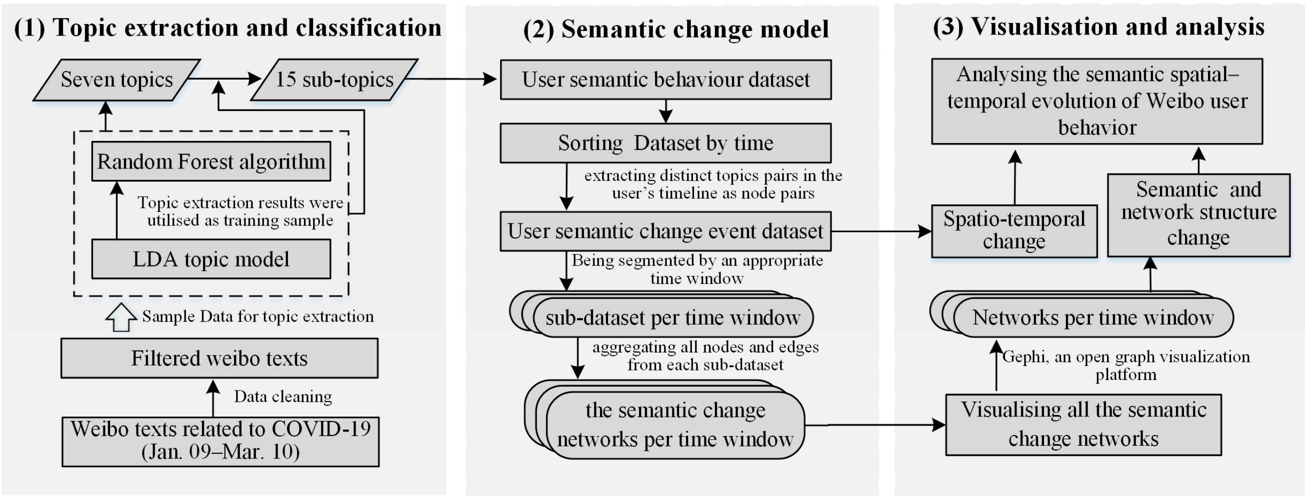

Figure 1. The framework for extracting the thematic evolution and temporal–spatial pattern of public behavior during disaster events.

3.2. Data and Pre-Processing

Using Sina Weibo APIs, we acquired Weibo data that contained keywords 'pneumonia' and/or 'coronavirus' related to COVID-19 from 00:00 on 9 January 2020, to 24:00 on 10 March 2020. The metadata information included user ID, time, text, and location information.

For pre-processing, we firstly filtered out the duplicated or/and very short (less than four words) Weibo texts. Then, noise data (e.g., hyperlinks, spaces, punctuation marks, hashtags, and @user mentions) were deleted. After text filtering, we were left with a canonical dataset of 3,427,933 Weibo texts posted by 986,217 unique Weibo users, including 197,118 texts with geographical coordinates generated by 98,507 unique Weibo users.

3.3. Topic Extraction and Classification Model

LDA and RF algorithm were combined to construct the topic extraction and classification model for hierarchically processing Weibo texts related to COVID-19, based on our previous work [**? ?**]. Firstly, LDA mode was used to process the Weibo sample to generalize the topics and assigned each Weibo sample to one topic, which were utilized to train the RF model. Then the trained RF model was applied to classify the entire Weibo dataset. Finally, the Weibo texts related to COVID-19 were generalized into seven topics. The topics of 'personal response', 'opinions and sentiments', and 'seeking help' were divided into 15 more detailed sub-topics by implementing a secondary classification. The details of topic and subtopics were shown in Figure **??**.

In the experiment, 8500 Weibo texts were randomly selected to generalize the seven topics by LDA model, which was implemented by the "Gensim" package in Python. Through repeated experiments, the optimal number of initial topics was set as 20. We assigned each Weibo text to the topic that it most closely resembled according to the probabilities in the document topic lists. Based on the topic-terminology lists, twenty topics were generalized into seven ("fifteen" in the secondary classification) by merging similar topics and discarding irrelevant topics.

The RF algorithm was implemented using a machine learning package called "scikit-learn" in Python. Based on the document topic lists, 7000 annotated Weibo texts were used as training samples and 1500 annotated Weibo texts were used as test sets. We used the out-of-bag (OOB) outputs to determine the number of classification trees (n estimators) at 400. In the secondary classification, we randomly selected 10,000 Weibo texts from the topics of 'personal response', 'opinions and sentiments', and 'seeking help' to generalize 15 sub-topics. There were 8000 training samples and 2000 test samples for training the RF model. The number of classification trees (n estimators) was 400.

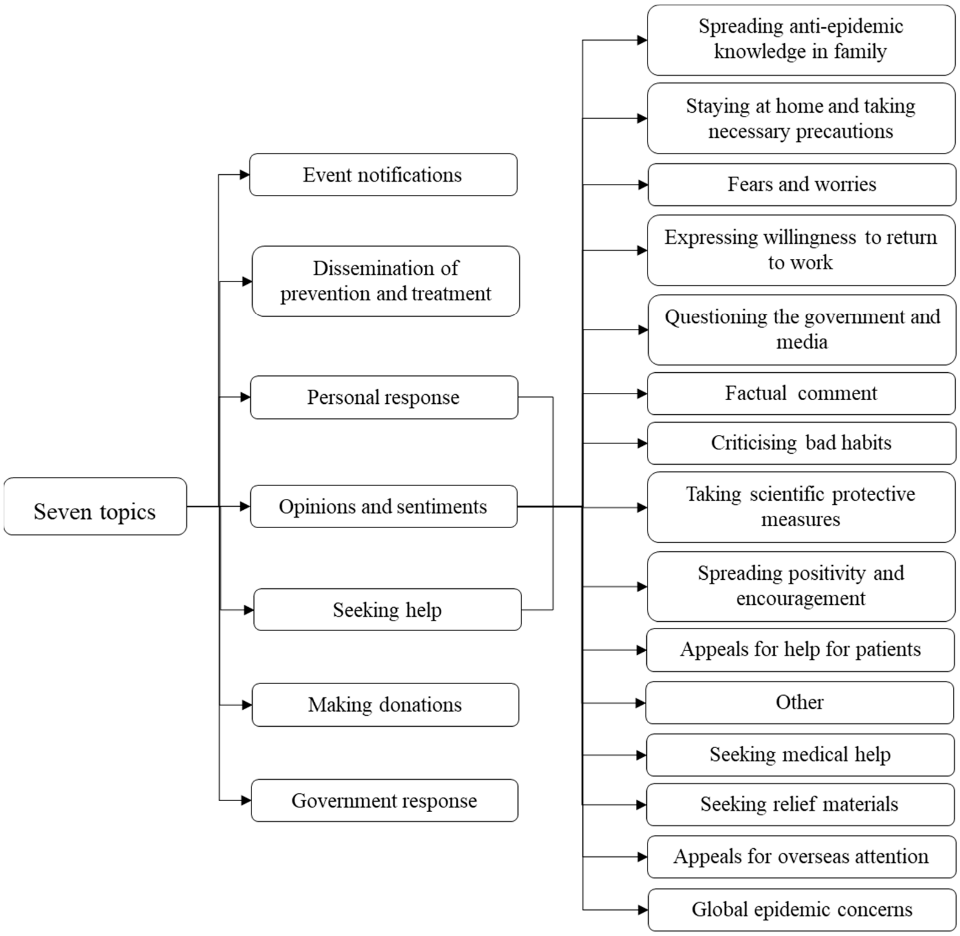

Figure 2. The classification of COVID-19-related topics and sub-topics.

3.4. Semantic Change Model

3.4.1. Network Measures

Based on complex networks, we proposed a semantic change network (see 3.4.2 for detailed definition), setting topics as nodes and topic changes as directed edges. The semantic change network allows us to quantitatively capture the semantic evolution process of social media users during COVID-19 by analyzing the complex network topology structure. We selected several important network metrics, including network size, degree, weighted average degree, and aggregation coefficient.

Network size: Network size is the number of topics (nodes) in the semantic change network.

Degree: the node degree is the number of edges connected to that node [?]. In undirected network, a node's degree is how many edges concatenate to it. In directed graphs, there are two measures of degree: in-degree and out-degree. The in-degree is the number of edges entering the node, while the out-degree is the number of edges originating from the node moving outward to other nodes. The sum of the in-degree and out-degree gives you the total degree for the node. In this study, the size of the nodes was determined by the weighted degree.

Average weighted degree: The edge weights represent the quantity of users whose Weibo texts changed semantically from one topic to another. The average weighted degree of the semantic change network reflects the semantic change frequency. Let k_i be the weighted degree of node i. The average weighted degree is defined by Equation (1):

$$\mathrm{K} = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{1}$$

Average clustering coefficient: the clustering coefficient is used to represent the degree of clustering between network nodes. The average clustering coefficient is specifically the tendency of nodes to generate triangles in networks [?]. Given a node i in network G, the probability that the k_i nodes connected to node i are also connected to each other is given by Equation (2):

$$C_i = \frac{2l_i}{k_i(k_i - 1)} \tag{2}$$

l_i represents the set of edges connected to node i and k_i represents the degree of the node i. The average clustering coefficient is calculated using Equation (3).

$$C = \frac{1}{N} \sum C_i \tag{3}$$

3.4.2. Semantic Change Network

Based on a complex network, we constructed a semantic change network to describe the process of semantic change in Weibo users' opinions about responses to an emergency. In the network, a node represents a topic, and the directed edge represents changing topics. When the use of topic A changing to B appears in one user's timeline, an edge is added from node A to node B. The edge weight indicates how many users generate thematic conversion from topic A to B. The detailed definitions are as follows.

User semantic behavior: a user's semantic behavior means that a user posts a Weibo message, in this case, related to disaster events. We characterize a user's semantic behavior as a tuple $r = (u, t, g, s)$, consisting of four facets describing the behavior: u is the user who posted the text (user ID); t is the time of the behavior (the time when Weibo text was posted); g is the spatial location of the behavior (geographical coordinates); and s is the thematic attributes of behavior (the topic of the Weibo text).

User semantic change event: the semantic change event of a user u is a transformation between two of u's semantic behaviors, denoted as $e_{ij} = \left((u, t_i, g_i, s_i), (u, t_j, g_j, s_j)\right)$, where $t_i < t_j$, $s_i \neq s_j$. When the topic published by a user changed from topic A to B, the user is considered to have produced a semantic change event.

Semantic change network: a semantic change network is a weighted, directed graph $TG = (T_V, T_E)$, as shown in Figure **??**. Given a set of users U and their Weibo texts within a limited time period $T = (t_T, t_T + \delta]$ for $\delta > 0$, T_V is the node set (or the topic set), including all semantic topics generated by users in time period T. The number of nodes represents the size of the network. The size of the nodes represents the number of edges coming into the node. T_E is the set of edges, denoted as $T_E = \left\{\left((u, t_i, g_i, s_i), (u, t_j, g_j, s_j)\right) : \forall u \in U, \forall t_i \in T \wedge \forall t_j \in T\right\}$. Each edge corresponds to a semantic behavior change event. The edge weight indicates how many times this semantic behavior occurs in the time period T. T_E includes all semantic change events produced by users in time period T.

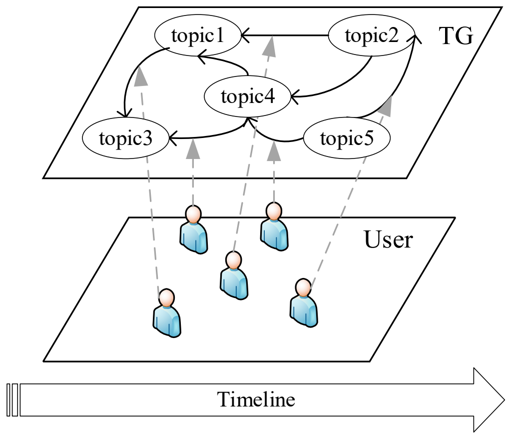

Figure 3. The structure of the semantic change network.

Based on the above definition, a semantic change network was constructed by the following steps: (1) we firstly sorted Weibo texts in chronological order and extracted topics from them. (2) For individual users, we acquired all the different topics that appeared in the user's timeline as vertices pairs, among which two adjacent topics in time constitute a topic pair with edge pointing to the most recently used topic. In this way, the user semantic change event dataset was generated. (3) To better visualize dynamic semantic changes in a long time series, all networks were constructed with a sliding time window. Each network is built with an n-day window length. The corresponding network at time point t_T is determined by semantic change events produced by users during time period $T = (t_T, t_T + n]$. The method of determining an appropriate time window n is presented in Section **??**. (4) Starting from the initial time, the user semantic change event dataset was segmented into several sub-datasets by the time window n. We then integrated all vertices and edges from each sub-dataset to build the semantic change networks. We used Gephi (Gephi Consortium, Paris, France), an open-source and free software for designing all kinds of graphs and networks, to visualize the semantic change network.

Fifteen sub-topic responses to COVID-19 were selected to construct the semantic change network. Because the sub-topic names were long, we codified the sub-topic names in the network construction, as shown in Table **??**.

Table 1. The sub-topic names and corresponding codes.

Sub-Topics	Code
Fears and worries	S1
Questioning the government and media	S2
Criticizing bad habits	S3
Factual comments	S4
Taking scientific protective measures	S5
Spreading positivity and encouragement	S6
Appeals for help for patients	S7
Expressing willingness to return to work	S8
Other	S9
Staying at home and taking necessary precautions	S10
Spreading anti-epidemic knowledge in family	S11
Seeking medical help	S12
Seeking relief materials	S13
Appeals for overseas attention	S14
Global epidemic concerns	S15

3.4.3. Determination of Time Window

A proper time window is an important parameter in constructing the semantic change networks. The time window is determined by both the number of users and the frequency of topics that change in different user sets. If the time window is too short, the semantic networks will be too sparse, resulting in insufficient nodes and edges to support the network structure. On the other hand, if the time window is too long, the adjacent semantic networks will be similar, which makes it difficult to detect changes in the network structure and leads to a waste of computing resources because of the large volume of each network.

In this study, two indicators, **the average number of semantic change events** and **the average increment of semantic change events**, were used to determine the time window. According to our definition, a semantic change event is a transition between two user topics. The topic pair (s_i, s_j) is the change pair that induces this semantic change event. **The average number of semantic change events** is the average number of edges in all semantic change networks in a given time window. **The average increment of semantic change events** is the proportion of new topic change pairs between two adjacent semantic networks.

When the time window $n \in [0.5, 10]$ is fixed, for the given time point d from 9 January 2020 to 10 March 2020, let $S_{d,n}$ denote all the topic pairs in $[d, d + n]$. We computed the increment of semantic change events between $S_{d,n}$ and $S_{d+n/2,n}$, defined as

$(S_{d+n/2,n} - S_{d,n})/S_{d,n}$. Then, the average of all the results was used to obtain **the average increment of semantic change events**. Similarly, we obtained **the average number of semantic change events** by calculating the average of all $S_{d,n}$.

As shown in Figure **??**, **the average increment of semantic change events** between two consecutive networks rapidly decreases as the time window increases, indicating that the larger the time window, the more similar the adjacent networks, which make it difficult to clearly reflect semantic changes. **The average number of semantic change events** increases with an increase in the time window in Figure **??**, showing that the larger the time window, the richer the semantic changes. Therefore, we used the elbow area in Figure **??**, where n = 2 days, as the optimal time window.

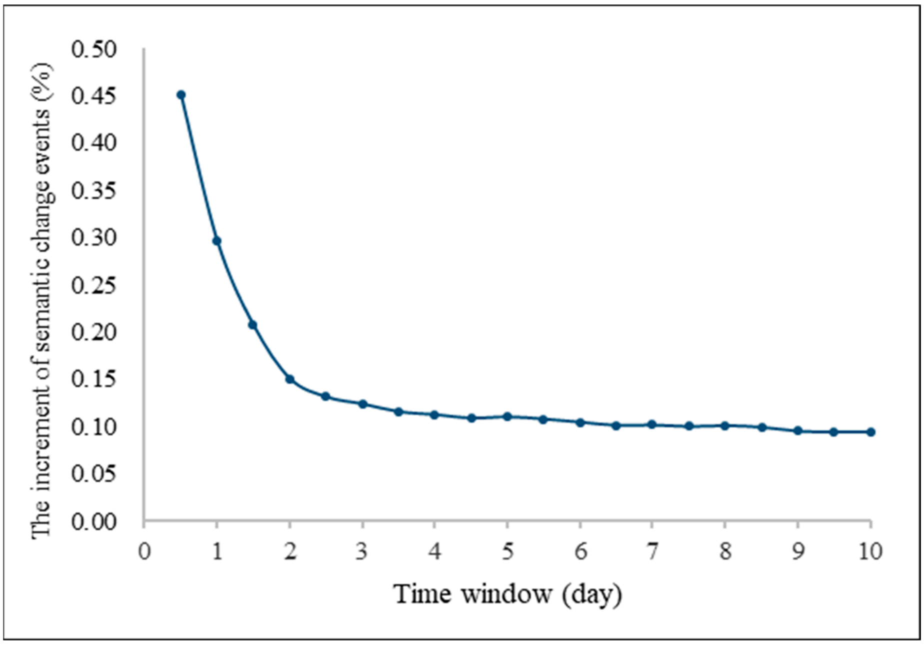

Figure 4. The average increment of semantic change events for different time windows.

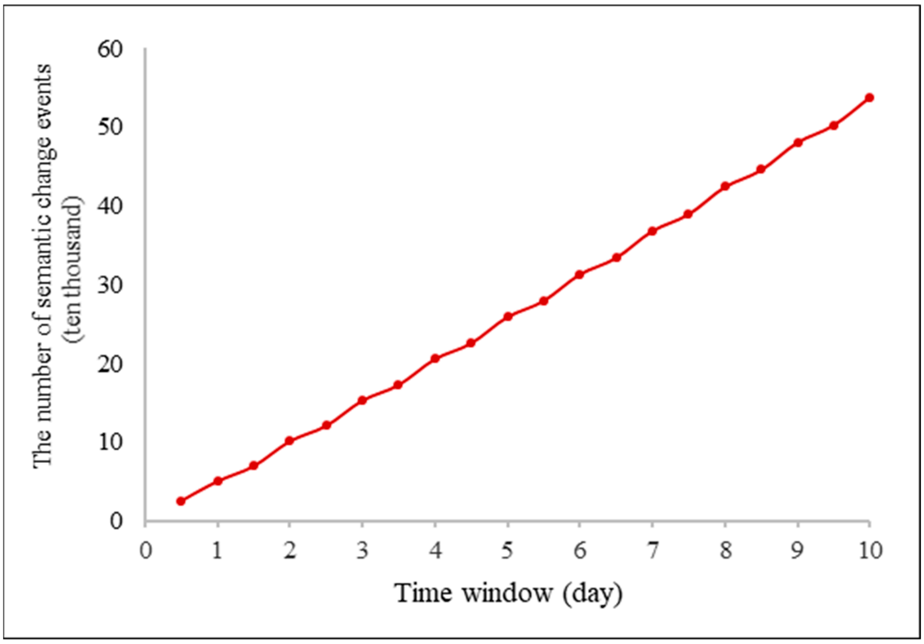

Figure 5. The average number of semantic change events for different time windows.

4. Results

4.1. Topic Description

COVID-19-related topics and sub-topics were counted and presented in Figure ??. 'Opinions and sentiments' accounted for 32.89% of all topics, followed by 'government response' and 'event notifications', which were 20.64% and 17.08%, respectively. 'Dissemination of prevention and treatment' and 'personal response' comprised 14.12% and 11.57%. 'Seeking help' and 'making donations' accounted for 1.47% and 2.23%.

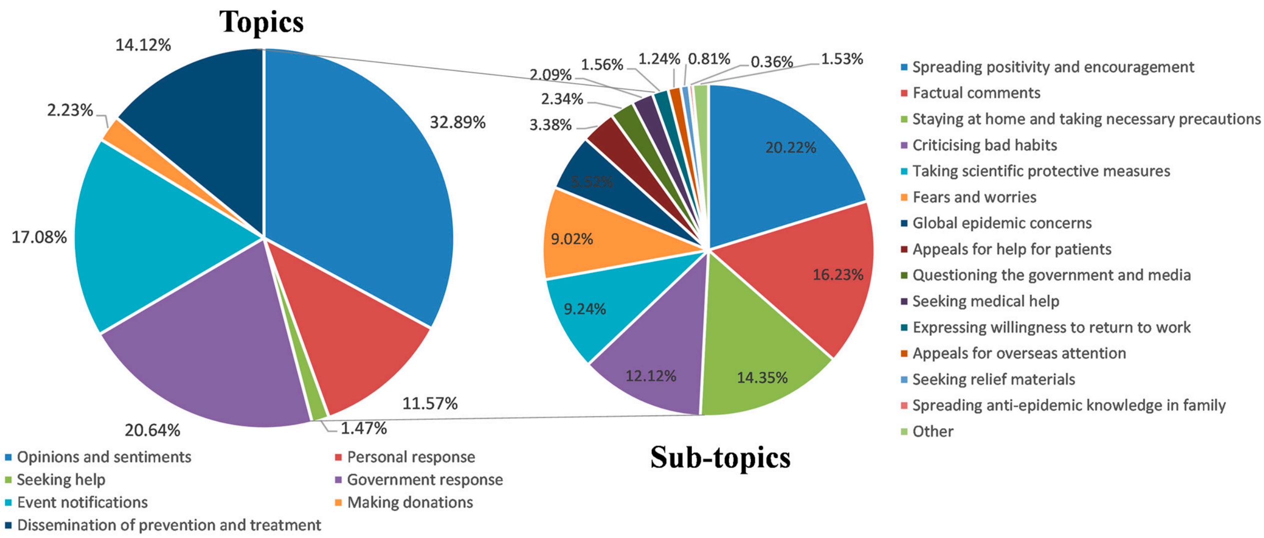

Figure 6. Classification of topics and sub-topics extracted from Weibo texts relevant COVID-19.

For sub-topics, the quantities of 'spreading positivity and encouragement', 'factual comments', and 'staying at home and taking necessary precautions' were the most, accounting for 20.22%, 16.23%, and 14.35%, respectively. The proportion of 'criticizing bad habits', 'taking scientific protective measures', and 'fears and worries' comprised 12.12%, 9.24%, and 9.02%. 'Global epidemic concerns' and 'appeals for help for patients' accounted for 5.52% and 3.38%, respectively. The remaining topics accounted for less than 3%.

4.2. Spatio-Temporal and Semantic Evolution Analyses

4.2.1. Visualization of the Semantic Change Network

???? showed the process of semantic changes of social media users' COVID-19-related topics from 9 January to 10 March 2020. The semantic network changes were analyzed in different time periods.

From 9 January to 19 January, there were few COVID-19-related Weibo posts, so the semantic change network was sparse. Over time, the number of network nodes (the number of topics) changed from the initial 5 to 11. The semantic change of topics mainly concentrated on 'S5-Taking scientific protective measures', 'S10-Staying at home and taking necessary precautions', 'S6-Spreading positivity and encouragement', 'S4-Factual comments', and 'S1-Fears and worries'.

From 20 January to 4 February, the amount of user semantic change events increased sharply on 20 January, peaked on 21 January, declined but fluctuated until 29 January, significantly increased on 30 January, and achieved a new peak on 1 February. It then smoothly fluctuated during 2-4 February. As the number of topics changed from 11 to 13, the semantic network became increasingly dense. The topics 'S7-Appeals for help for patients' and 'S3-Criticising bad habits' appeared. In the early stage, public opinion mainly changed between 'S10-Staying at home and taking necessary precautions', 'S6- Spreading positivity and encouragement', 'S5-Taking scientific protective measures', 'S4-Factual comments', and 'S1-Fears and worries'. Users' topics then shifted to 'S3-Criticising bad habits', before finally moving to 'S10-Staying at home and taking necessary precautions', 'S5-Taking scientific protective measures', 'S6-Spreading positivity and encouragement', and 'S4-Factual comments'.

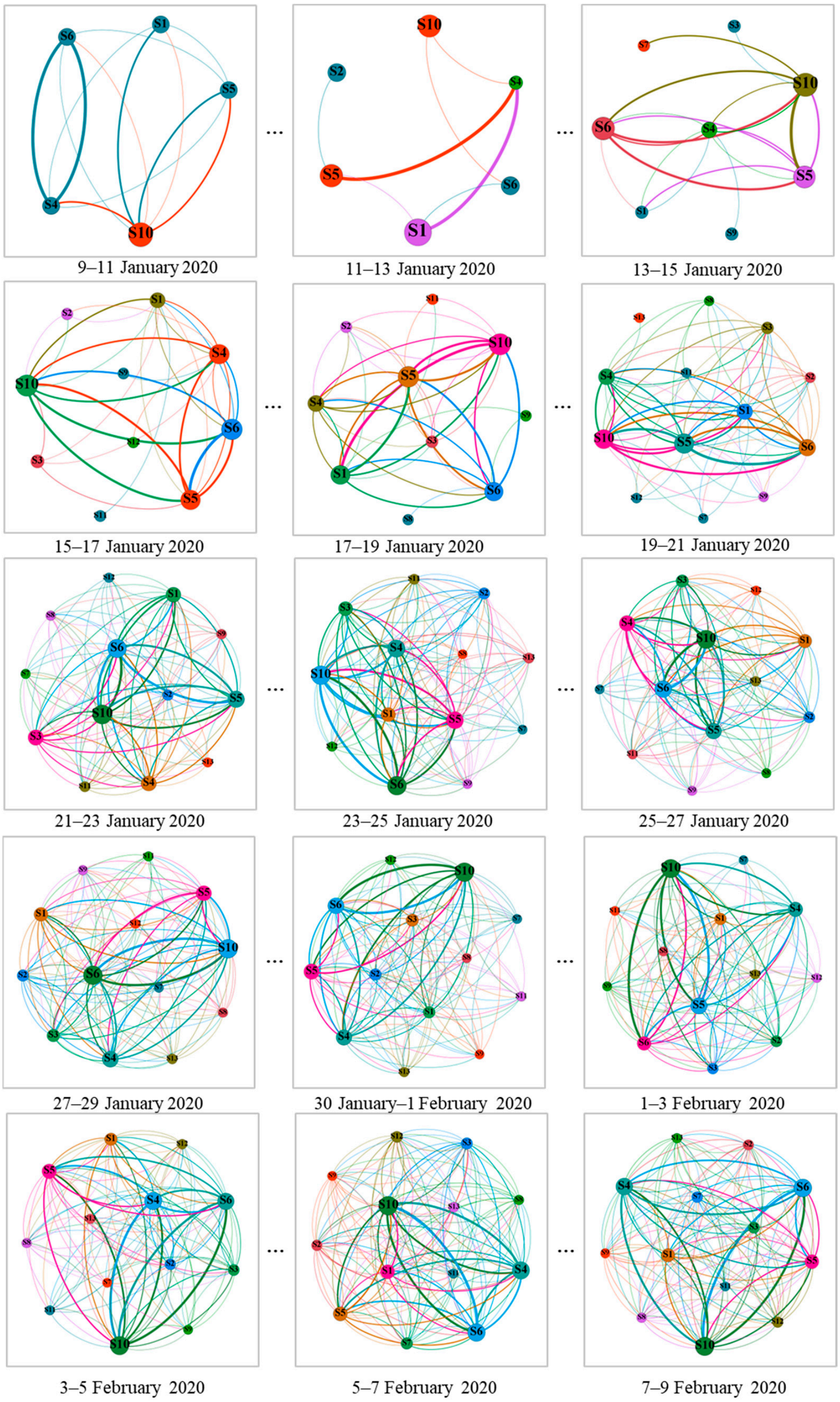

Figure 7. Semantic change network (9 January–9 February 2020) (S1: Fears and worries; S2: Questioning the government and media; S3: Criticizing bad habits; S4: Factual comments; S5: Taking scientific protective measures; S6: Spreading positivity and encouragement; S7: Appeals for help for patients; S8: Expressing willingness to return to work; S9: Other; S10: Staying at home and taking necessary precautions; S11: Spreading anti-epidemic knowledge in the family; S12: Seeking medical help; S13: Seeking relief materials; S14: Appeals for overseas attention; S15: Global epidemic concerns).

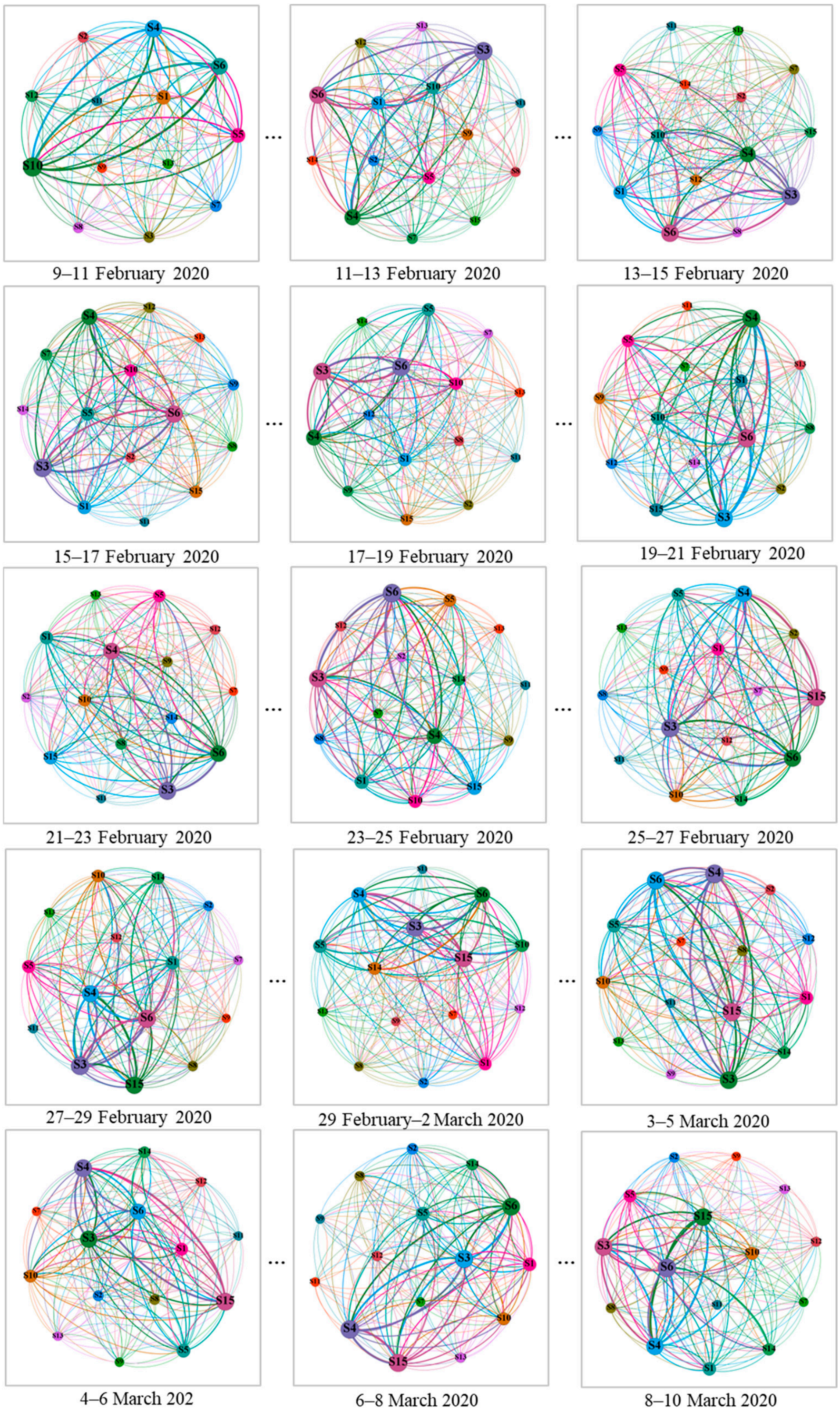

Figure 8. Semantic change network (8 February–10 March 2020) (S1: Fears and worries; S2: Questioning the government and media; S3: Criticizing bad habits; S4: Factual comments; S5: Taking scientific protective measures; S6: Spreading positivity and encouragement; S7: Appeals for help for patients; S8: Expressing willingness to return to work; S9: Other; S10: Staying at home and taking necessary precautions; S11: Spreading anti-epidemic knowledge in the family; S12: Seeking medical help; S13: Seeking relief materials; S14: Appeals for overseas attention; S15: Global epidemic concerns).

From 5 February to 24 February, public opinion began to change to 'S1-Fears and worries'. The convergence of 'S5-Taking scientific protective measures' decreased. On 9 February, the topic change mainly concentrated on 'S10-Staying at home and taking necessary precautions', 'S6-Spreading positivity and encouragement', and 'S4-Factual comments'. 'S5-Taking scientific protective measures' and 'S1-Fears and worries' were second. On 10 February, the number of topics increased from 13 to 15. The topic 'S14-Appeals for overseas attention' and 'S15-Global epidemic concerns' appeared. The topic changes began to gather around 'S3- Criticizing bad habits' and 'S15-Global epidemic concerns'. 'S10-Staying at home and taking necessary precautions' and 'S5-Taking scientific protective measures' decreased. A pattern gradually formed whereby 'S3-Condemning bad habits', 'S6-Spreading positivity and encouragement', and 'S4-Factual comments' were the main cluster, followed by 'S5-Taking scientific protective measures', 'S10-Staying at home and taking necessary precautions', and 'S15-Global epidemic concerns'.

From 25 February to 10 March, the semantic changes continued to converge toward 'S15-Global epidemic concerns'. 'S10-Staying at home and taking necessary precautions' and 'S5-Taking scientific protective measures' continued to decline, while the convergence on 'S14-Appeals for overseas attention' rose. Finally, changes mainly concerned 'S15-Global epidemic concerns', 'S6-Spreading positivity and encouragement', 'S4-Factual comments', and 'S3-Condemning bad habits', followed by 'S5-Taking scientific protective measures', 'S10-Staying at home and taking necessary precautions', 'S1-Fears and worries', and 'S14-Appeals for overseas attention'.

Based on the changes in network structure (Figure **??**), we summarized the characteristics of the user semantic change pattern during the COVID-19 outbreak. As shown in Figure **??**a,b, the change pattern of users' topics was from expansive to stable. The average weighted degree (Figure **??**a) fluctuated with time and increased significantly with key social events, indicating that the rate of user topic change increased. The average aggregation coefficient increased significantly in the early stages and then stabilized. The higher average aggregation coefficient indicates that the users' topics changed more easily to form a closed loop of repeated changes among several topics. It can be seen that in the prophase of the COVID-19, social media users' topics changed frequently. With the development of the event, the changes in user topics tended to stabilize and were concentrated in a few topics. Figure **??**c,d show that the semantic change network had clear expansibility in the early stage. The number of new semantic change events increased significantly at the beginning of the COVID-19 outbreak, tended to be stable during the follow-up, and surged at several key social events.

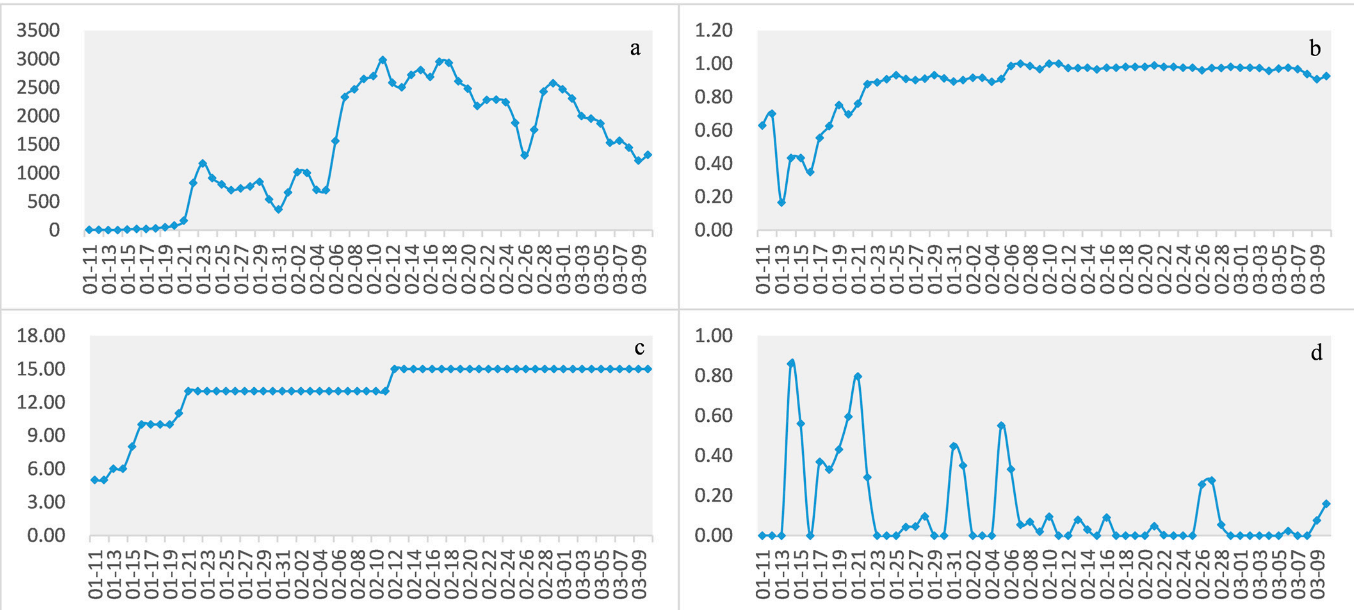

Figure 9. The metric values of semantic network: (**a**) average weighted degree; (**b**) average clustering coefficient; (**c**) network size; (**d**) average increment of semantic change events.

4.2.2. Spatial Analysis of the Semantic Change Network

Statistical analysis (Figure **??**) revealed a total of 1,544,109 user semantic behaviors generated by 118,137 users, and 324,900 user semantic change events (including 24,328 events with geographical coordinates). The top three semantic change events were ('S6-Spreading positivity and encouragement' to 'S3-Condemning bad habits'), ('S6-Spreading positivity and encouragement' to 'S4-Factual comments'), and ('S10-Staying at home and taking necessary precautions' to 'S6-Spreading positivity and encouragement'), with 21,022, 20,929, and 20,644 changes, respectively.

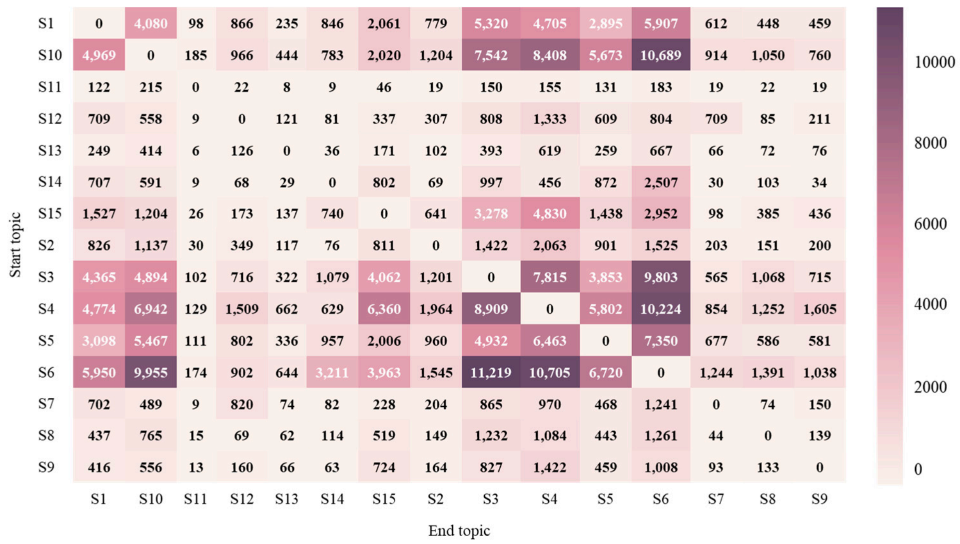

Figure 10. Heat map of user semantic change events.

The spatial distribution of the semantic change events is shown in Figure **??**. The most frequent changes in user topics focused on the Beijing–Tianjin–Hebei region, Yangtze River Delta region, and the provinces of Hubei, Sichuan, and Guangdong. Figure **??** shows a more detailed analysis of the above five regions. In the Beijing–Tianjin–Hebei region, user semantic changes were mainly centered at Beijing, Tianjin, and Shijiazhuang. In the Yangtze River Delta region, Shanghai and Hangzhou were the cities with the largest number of topic changes, followed by Ningbo, Nanjing, Hefei, and Suzhou. The semantic change events in Hubei Province were mainly distributed in Wuhan, Huanggang, Xiaogan, Xiangyang, and Yichang. Guangzhou, Shenzhen, and their surrounding cities were the main cities of semantic change events in the Guangdong Province. Changes in the Sichuan Province were mainly concentrated in Chengdu, Mianyang, and Nanchong.

We then constructed semantic change networks for the five areas identified (Figure **??**). The semantic changes in the Beijing–Tianjin–Hebei region mainly focused on 'S6-Spreading positivity and encouragement', 'S10-Staying at home and taking necessary precautions', 'S3-Condemning bad habits', 'S4-Factual comments', 'S1-Fears and worries', 'S5-Taking scientific protective measures', and 'S15-Global epidemic concerns'. The Yangtze River Delta was dominated by 'S6-Spreading positivity and encouragement', 'S10-Staying at home and taking necessary precautions', 'S3-Condemning bad habits', 'S4-Factual comments', 'S5-Taking scientific protective measures', and 'S1-Fears and worries'. The topics of Hubei Province mainly changed among 'S6-Spreading positivity and encouragement', 'S10-Staying at home and taking necessary precautions', 'S3-Condemning bad habits', 'S4-Factual comments', 'S1-Fears and worries', and 'S5-Taking scientific protective measures'. The semantic changes in the Guangdong Province mainly focused on 'S10-Staying at home and taking necessary precautions', 'S6-Spreading positivity and encouragement', 'S3-Condemning bad habits', 'S4-Factual comments', 'S5-Taking scientific protective measures', 'S1-Fears and worries', and 'S15-Global epidemic concerns'. The Sichuan Province was dominated by 'S6-Spreading positivity and encouragement', 'S10-Staying at home and taking necessary precautions', 'S3-Condemning bad habits', 'S4-Factual comments', and 'S5-Taking scientific protective measures'.

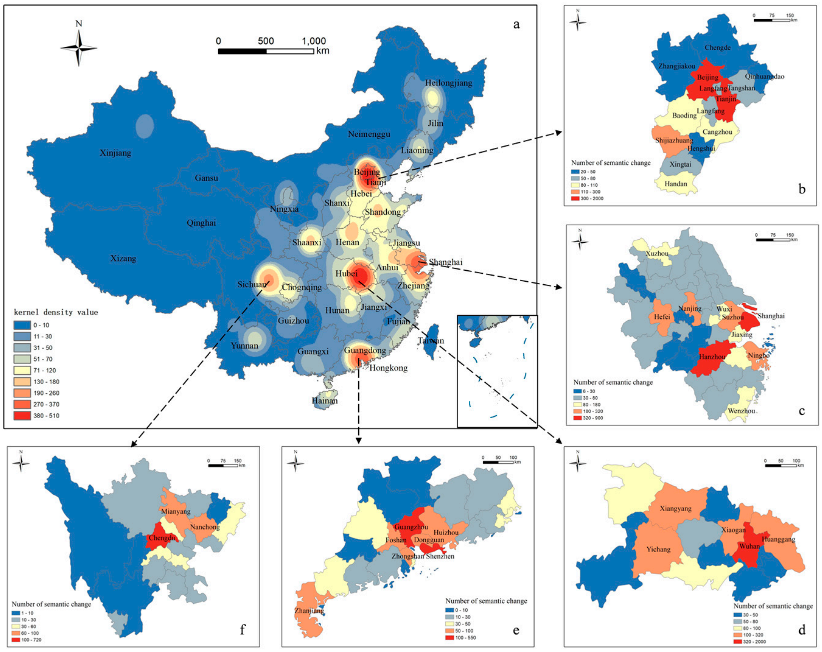

Figure 11. The spatial distribution of COVID-19-related semantic change events. Sub-figures (**a**–**f**) show the kernel density of semantic change events in: (**a**) China; (**b**) Beijing–Tianjin–Hebei region; (**c**) Yangtze River Delta region; (**d**) Hubei Province; (**e**) Guangdong Province; (**f**) Sichuan Province.

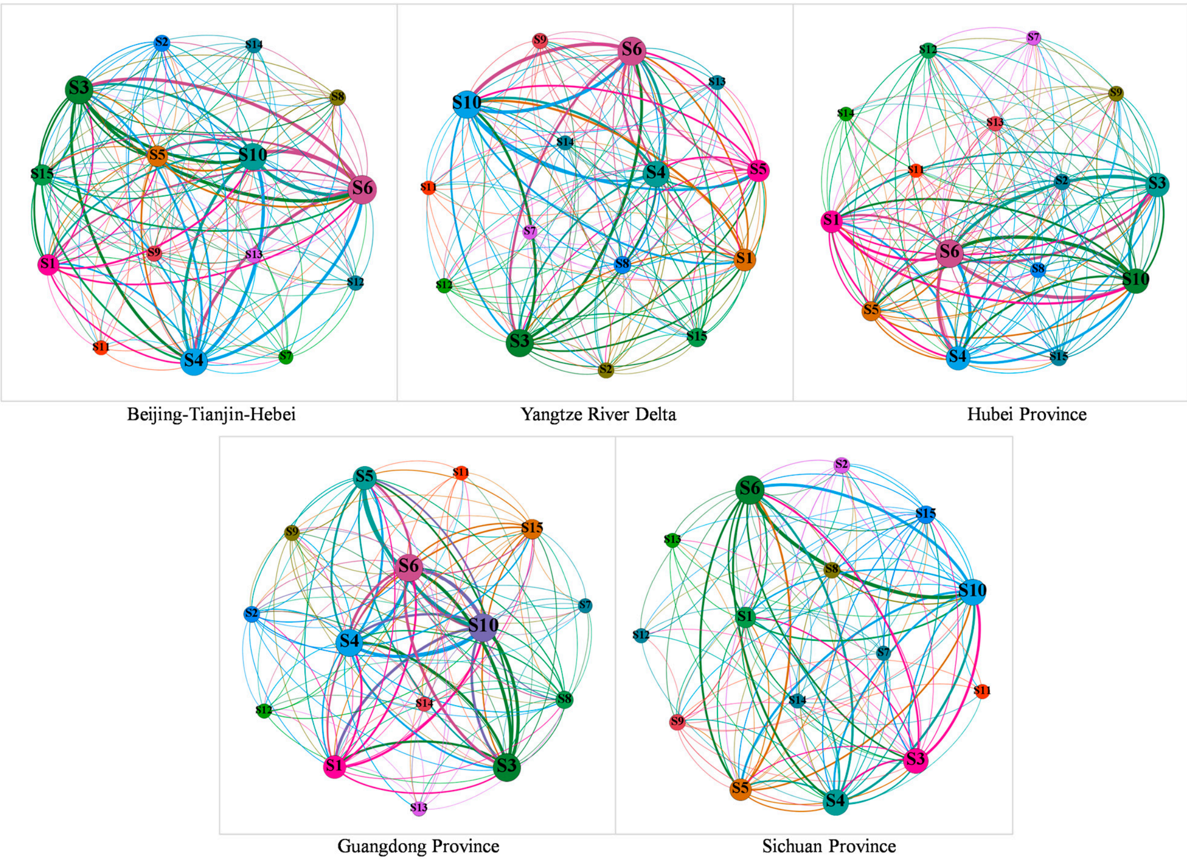

Figure 12. The semantic change networks in five areas.

5. Discussion

By integrating topic extraction, complex networks, and GIS, this framework intuitively shows the spatial, temporal, and semantic evolution of the topics of Weibo users' responses to COVID-19. During the prophase of COVID-19 outbreak, Weibo users' semantic change networks had obvious expansibility. The increases in the average weighting degree and average clustering coefficient reflected the frequent and repeated changes of social media users between various topics, indicating the anxiety felt by users who were trying to understand all aspects and details of events. Over time, the semantic change of social media users' topics gradually became focused on a few topics, and the network indicators stabilized. These findings are consistent with those of previous studies on different disaster events. He [?] found that social media users were more likely to switch topics in their tweets immediately following an event. Zhang [?] found that public attention expands rapidly in the early stages of disaster events and then gradually diminishes, until finally the public pays no more attention to the disaster.

From a methodological perspective, the previous studies of semantic dynamics have focused on analyzing the temporal—spatial changes of the number or subjects of topics [? ? ?] or improving the algorithm of topic models from the perspective of natural language processing [?]. In this paper, the complex semantic evolution process of Weibo users is mapped to a complex network, and changes in user semantics are transformed into changes in the network structure. Considering the spatiotemporal characteristics of social media data, we defined the semantic change network systematically, and described the dynamic temporal, spatial, and semantic characteristics of public behavior during the COVID-19 outbreak.

From the view of semantics, this paper provides the analysis of public topical evolution during the COVID-19 in China. From 9 January 2020 to 10 March 2020, the overall trend of China's public opinion is rational and positive. According to the different focus of topics, the user semantic evolution can be divided into four stages that correspond to the development of the COVID-19 event. In the first stage, the topics were fewer, and the network was sparse. The semantic changes mainly converged on S5. In the second stage, the number of related Weibo texts and semantic change events increased sharply. The main changes in this stage focused on S10, S6, S4, and S5. In the third stage, the domestic epidemic situation in China began to improve and the foreign epidemic situation began to ferment. The topic changes mainly focused on S3, S6, and S4. After 10 February, S14 and S15 appeared. The fourth stage was when the epidemic situation in China had stabilized while the confirmed cases number in the world presented a sustained rise. In this stage, the main semantic changes were between S15, S6, S4, and S3. Our results are consistent with previous work, in which public sentiment was strongly associated with exterior elements, for example, the influence of official media, significant social events, violent weather, and public holidays [? ?].

In terms of spatial distribution, the Beijing–Tianjin–Hebei region, Yangtze River Delta region, Hubei Province, Sichuan Province, and Guangdong Province were the areas where user topics changed most frequently, mainly changing among S6, S10, S3, and S4. However, the focus of public opinion varied across regions. Changes between S1, S5, and S15 were quite frequent in the Beijing–Tianjin–Hebei region and Guangdong Province, while the topic changes of S1 and S5 were more prominent in the Yangtze River Delta and Hubei Province, and only S5 was prominent in the Sichuan Province.

Nevertheless, this study has some limitations. First, the demographics of social media users show an obvious younger age structure [?], which makes it unrepresentative of the general population. Moreover, not all Weibo data include geographical coordinates, so the spatial analysis could not be applied to all the data. Second, due to confidentiality, this study ignores the geographical–social relationship among users, which is important to analyze the spatiotemporal characteristics of semantic change. Future research should focus on the geographic bias of social media and explore the spatial, temporal, and semantic evolution pattern of public opinion on social networks. Third, in normal circumstances (nonemergency situation), The communication topics on social media (e.g., Twitter) are related to daily life routines, human activities, and interests, such as school, work, sports, dating, wearing, music, and food [**? ?**]. In emergencies situation, the public behaviors related to disasters emerge in large numbers in a short time. The social–geographical differences of public behavior between normal and abnormal situation deserve for further research.

6. Conclusions

From a methodological perspective to describe the dynamics of social media users' behavioral responses to disaster events in time, space, and semantics, this paper proposes a novel analysis framework of social media users, consisting of topic extraction, user semantic evolution model, and GIS spatial analysis. The user semantic evolution model systematically defined the semantic changes of social media users; mapped the semantic evolution of user topics into geographic space–time; and revealed the dynamic spatial, temporal, and semantic evolution characteristics of user topics during a disaster event. Taking the case of COVID-19 event, this study verified the practicability and effectiveness of the framework based on Sina Weibo data. The proposed framework is with the characteristics of generality and applicable for other social media data. It provides a methodological basis for comprehensive mining and analysis of social media user behavior during disaster events. In emergency situations, this framework could improve situational assessment, assist decision makers to better comprehend public opinion, and support analysts in allocating resources appropriately.

This research can be extended in several directions. First, future research can deepen the semantic change analysis from the users' perspective. By observing the semantic change of individual users and combining them with the geographical and social relations among them, the semantic changes of group users can be collected to explore the group semantic change pattern during disaster events. Second, for the results of analysis and visualization, it is necessary to develop an interactive user semantic change geographic visualization analysis platform that can visually display and interactively analyze the spatial–temporal distribution of user semantic change, user geographic social networks, user semantic change patterns, and statistical information in disaster events. Third, we assigned each Weibo text to the topic it most closely resembled according to the probabilities in the document topic lists generate by the LDA model in this study. Actually, a Weibo post may contain multiple topics. The multiple topics of social media texts is also an interesting and meaningful issue worth in-depth study.

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

Funding: This research was funded by the Chinese Academy of Sciences (No. ZDRW-XH-2021-3) and the Construction Project of China Knowledge Center for Engineering Sciences and Technology (No. CKCEST-2022-1-41).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Miller, H.J.; Goodchild, M.F. Data-driven geography. *Geo J.* **2015**, *80*, 449–461. [CrossRef]
2. Steiger, E.; Resch, B.; Zipf, A. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1694–1716. [CrossRef]
3. Koylu, C. Modeling and visualizing semantic and spatio-temporal evolution of topics in interpersonal communication on Twitter. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 805–832. [CrossRef]
4. Dunkel, A.; Andrienko, G.; Andrienko, N.; Burghardt, D.; Hauthal, E.; Purves, R. A conceptual framework for studying collective reactions to events in location-based social media. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 780–804. [CrossRef]
5. Koylu, C.; Larson, R.; Dietrich, B.J.; Lee, K.-P. CarSenToGram: Geovisual text analytics for exploring spatiotemporal variation in public discourse on Twitter. *Cartogr. Geogr. Inf. Sci.* **2019**, *46*, 57–71. [CrossRef]
6. Wang, Y.; Wang, T.; Ye, X.; Zhu, J.; Lee, J. Using social media for emergency response and urban sustainability: A case study of the 2012 Beijing rainstorm. *Sustainability* **2016**, *8*, 25. [CrossRef]
7. Resch, B.; Usländer, F.; Havas, C. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 362–376. [CrossRef]
8. Han, X.; Wang, J. Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 185. [CrossRef]
9. Han, X.; Wang, J.; Zhang, M.; Wang, X. Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2788. [CrossRef]
10. Dahal, B.; Kumar, S.A.; Li, Z. Topic modeling and sentiment analysis of global climate change tweets. *Soc. Netw. Anal. Min.* **2019**, *9*, 24. [CrossRef]
11. Jiang, Y.; Li, Z.; Cutter, S.L. Social network, activity space, sentiment, and evacuation: What can social media tell us? *Ann. Am. Assoc. Geogr.* **2019**, *109*, 1795–1810. [CrossRef]
12. Zou, L.; Lam, N.S.; Shams, S.; Cai, H.; Meyer, M.A.; Yang, S.; Lee, K.; Park, S.-J.; Reams, M.A. Social and geographical disparities in Twitter use during Hurricane Harvey. *Int. J. Digit. Earth* **2019**, *12*, 1300–1318. [CrossRef]
13. Kryvasheyeu, Y.; Chen, H.; Obradovich, N.; Moro, E.; Van Hentenryck, P.; Fowler, J.; Cebrian, M. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2016**, *2*, e1500779. [CrossRef] [PubMed]
14. Yuan, Y.; Lu, Y.; Chow, T.E.; Ye, C.; Alyaqout, A.; Liu, Y. The Missing Parts from Social Media–Enabled Smart Cities: Who, Where, When, and What? *Ann. Am. Assoc. Geogr.* **2020**, *110*, 462–475. [CrossRef]
15. Wang, Z.; Ye, X. Social media analytics for natural disaster management. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 49–72. [CrossRef]
16. Zou, L.; Lam, N.S.; Cai, H.; Qiang, Y. Mining Twitter Data for Improved Understanding of Disaster Resilience. *Ann. Am. Assoc. Geogr.* **2018**, *108*, 1422–1441. [CrossRef]
17. Huang, Q.; Xiao, Y. Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1549–1568. [CrossRef]
18. Wang, J.; Hu, Y.; Joseph, K. NeuroTPR: A neuro-net toponym recognition model for extracting locations from social media messages. *Trans. GIS* **2020**, *24*, 719–735. [CrossRef]
19. Scheele, C.; Yu, M.; Huang, Q. Geographic context-aware text mining: Enhance social media message classification for situational awareness by integrating spatial and temporal features. *Int. J. Digit. Earth* **2021**, *14*, 1721–1743. [CrossRef]
20. Chen, Y.; Li, Y.; Wang, Z.; Quintero, A.J.; Yang, C.; Ji, W. Rapid Perception of Public Opinion in Emergency Events through Social Media. *Nat. Hazards Rev.* **2022**, *23*, 04021066. [CrossRef]
21. Gruebner, O.; Lowe, S.R.; Sykora, M.; Shankardass, K.; Subramanian, S.; Galea, S. Spatio-temporal distribution of negative emotions in New York City after a natural disaster as seen in social media. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2275. [CrossRef]
22. Gongora-Svartzman, G.; Ramirez-Marquez, J.E. Social Cohesion: Mitigating Societal Risk in Case Studies of Digital Media in Hurricanes Harvey, Irma, and Maria. *Risk Anal.* 2021; *online ahead of print*. [CrossRef]
23. Shi, W.-Z.; Zeng, F.; Zhang, A.; Tong, C.; Shen, X.; Liu, Z.; Shi, Z. Online public opinion during the first epidemic wave of COVID-19 in China based on Weibo data. *Humanit. Soc. Sci. Commun.* **2022**, *9*, 159. [CrossRef]
24. Zong, Q.; Yang, S.; Ying, C.; Shen, H. Behavior of Social Media Users in Disaster Area under the Outburst Disasters:A Content Analysis and Longitudinal Study of Explosion in Tianjin 12~(th) August 2015. *J. Inf. Resour. Manag.* **2017**, *7*, 13–19.
25. Burnap, P.; Williams, M.L.; Sloan, L.; Rana, O.; Housley, W.; Edwards, A.; Knight, V.; Procter, R.; Voss, A. Tweeting the terror: Modelling the social media reaction to the Woolwich terrorist attack. *Soc. Netw. Anal. Min.* **2014**, *4*, 206. [CrossRef]
26. Ye, X.; Li, S.; Yang, X.; Qin, C. Use of social media for the detection and analysis of infectious diseases in China. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 156. [CrossRef]
27. Kim, D.; Oh, A. Topic Chains for Understanding a News Corpus. In Proceedings of the Computational Linguistics and Intelligent Text Processing, Tokyo, Japan, 20–26 February 2011; Springer: Berlin/Heidelberg, Germany, 2011.
28. Blei, D.M.; Lafferty, J.D. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; Association for Computing Machinery: New York, NY, USA, 2006; pp. 233–240.
29. Ahmed, A.; Xing, E.P. Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*; AUAI Press: Catalina Island, CA, USA, 2010; pp. 20–29.

30. Han, J.; Huang, Y.; Kumar, K.; Bhattacharya, S. Time-Varying Dynamic Topic Model: A Better Tool for Mining Microblogs at a Global Level. *J. Glob. Inf. Manage.* **2018**, *26*, 104–119. [CrossRef]
31. He, X.; Lin, Y.-R. Measuring and monitoring collective attention during shocking events. *EPJ Data Sci.* **2017**, *6*, 30. [CrossRef]
32. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world'networks. *Nature* **1998**, *393*, 440–442. [CrossRef]
33. Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef]
34. Golbeck, J. Chapter 3-Network Structure and Measures [M]//GOLBECK. In *Analyzing the Social Web*; Morgan Kaufmann: Boston, MA, USA, 2013; pp. 25–44.
35. Xiao, Y.; Wang, Z.; Li, Q.; Li, T. Dynamic model of information diffusion based on multidimensional complex network space and social game. *Phys. A Stat. Mech. Its Appl.* **2019**, *521*, 578–590. [CrossRef]
36. Zhou, Y.; Guan, X.; Zhang, Z.; Zhang, B. Predicting the tendency of topic discussion on the online social networks using a dynamic probability model. In Proceedings of the Hypertext 2008 Workshop on Collaboration and Collective Intelligence, Pittsburgh, PA, USA, 19–21 June 2008; pp. 7–11.
37. Centola, D. The spread of behavior in an online social network experiment. *Science* **2010**, *329*, 1194–1197. [CrossRef] [PubMed]
38. Zheng, M.; Lü, L.; Zhao, M. Spreading in online social networks: The role of social reinforcement. *Phys. Rev. E* **2013**, *88*, 012818. [CrossRef]
39. Suh, B.; Hong, L.; Pirolli, P.; Chi, E.H. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In Proceedings of the 2010 IEEE Second International Conference on Social Computing, Minneapolis, MN, USA, 20–22 August 2010; pp. 177–184.
40. Su, Q.; Huang, J.; Zhao, X. An information propagation model considering incomplete reading behavior in microblog. *Phys. A Stat. Mech. Its Appl.* **2015**, *419*, 55–63. [CrossRef]
41. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef] [PubMed]
42. Hartman, R.; Seyednezhad, S.; Pinheiro, D.; Faustino, J.; Menezes, R. Entropy in network community as an indicator of language structure in emoji usage: A Twitter study across various thematic datasets. In Proceedings of the International Conference on Complex Networks and Their Applications, Cambridge, UK, 11–13 December 2018; pp. 328–337.
43. Zhang, S.; Xu, K.; Li, H. Measurement and analysis of information propagation in online social networks like microblog. *J. Xi'an Jiaotong Univ.* **2013**, *47*, 124–130.
44. Cao, X.; MacNaughton, P.; Deng, Z.; Yin, J.; Zhang, X.; Allen, J.G. Using twitter to better understand the spatiotemporal patterns of public sentiment: A case study in Massachusetts, USA. *Int. J. Environ. Res. Public Health* **2018**, *15*, 250. [CrossRef]