

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Community Evolutional Network for Situation Awareness Using Social Media

XIAOKANG FU¹, YANDONG WANG^{1,2}, MENG MENG LI¹, MINGXUAN DOU¹, MENG LING QIAO¹ AND KAI HU^{3,4}

¹State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China (e-mail: whistle_ant@whu.edu.cn)

²Collaborative Innovation Center for Geospatial Information Technology, Wuhan, China

³Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, Jiangsu, China

⁴School of Internet of Things, Jiangnan University, Wuxi 214122, China

Corresponding author: Yandong Wang (e-mail: ydwang@whu.edu.cn).

This work was supported in part by the National Key Research Program of China under Grant 2016YFB0501403, the National Natural Science Foundation of China under Grant 41271399, and China Special Fund for Surveying, Mapping and Geoinformation Research in the Public Interest under Grant 201512015.

ABSTRACT

Social media is important for situational awareness during a disaster. During a disaster, the situation of emergence often changes over time and hence the topics of social media messages generated by social media users also change accordingly. Few studies quantitatively describe the topic evolution of social media during a disaster and the corresponding relationship between topic evolution and disaster process. We address this problem using co-word network analysis and present a new method based on the community evolution of the co-word network to analyze topic evolution over time in social media. The method uses communities of the co-word network in social media to represent topics. Based on the theory of community evolution, a community evolutional network is proposed to support and quantify the evolution of the topics. We implemented the proposed method in a case study, “July 2012 Beijing flood” using the Sina Weibo dataset. Results show that our method can well quantify the evolution process of topics and validate the effectiveness of our method in real-world applications. The method can facilitate the understanding of public expression dynamics during a disaster and be used to reveal the process and stages of a disaster.

INDEX TERMS co-word network, community evolution, topic evolution, situational awareness

I. INTRODUCTION

DURING an emergency, government and nongovernment relief agencies must collect and understand related crisis information as much as possible, which is crucial for performing rescue work. To collect and understand related crisis information (i.e., to know what is happening in the affected communities during an event) is also known as situational awareness [1].

Previously, situational awareness has been obtained by social survey methods such as telephone calls, direct observations, or personal interviews. Such social surveys at the city level require years of dedicated resource investment to be successful [1]. With the rapid development of social network platforms such as Twitter, Facebook, and Flickr, many researchers have begun to use social media data to solve intractable problems across many domains [2]. Social media can be used as a real-time human sensor during

an event [3], which is an important role for situational awareness. A representative example is the Twitter-based disaster event warning system called Emergency Situation Awareness, which is developed by the Australian Academy of Science using natural language processing and text mining techniques [4]. On January 26, 2013, the system successfully detected a tornado attack in Queensland. It is clear that the application of social media in context-awareness has become more typical.

The use of social media for situation awareness is typically based on semantic mining and text categorization. Researchers often use semantic mining techniques to extract positive or negative perspectives on emergencies from social media texts. For example, by classifying Weibo messages into positive and negative manually, Wang et al. constructed two indicators that reflect air quality, thereby better estimating air quality and monitoring hazy weather [2]. Text cate-

gorization can be used to filter social media data or extract information directly related to emergency. For example, Li et al. conducted a text classifier based on deep convolutional neural networks (CNN) to filter microblogs that are the true and real-time descriptions of rainfall, followed by visualizing the filtered microblogs as heat maps to demonstrate the effectiveness of using social media to monitor heavy rainfall events [5]. Kongthon et al. (2012) proposed a series of social media text categories in disasters during floods in Thailand, including situation updates, disaster relief efforts, request assistance, disaster relief coordination, critical government, and emotionally relevant messages [6]. These categories are instructive for the classification of social media data in disasters.

The situation of a disaster often changes over time and hence the topics of social media messages generated by social media users. Researchers indicated that the evolution of disaster-related topics in time in social media messages could be coordinated with the process and disaster stage. For example, Huang et al. (2015) presented a coding schema for categorizing social media messages into different topics within different disaster stages (mitigation, preparedness, emergency response, and recovery). They analyzed these topics' trends over time and their distribution in space, which can offer a better insight into the complex environment in a time of crisis [1]. Such studies divide social media messages directly into topics that correspond to the disaster stages. However, some topics may appear in several stages and the stages they belong to are difficult to distinguish. Furthermore, researchers classified social media messages in more detail and then analyzed the disaster stage through time series analysis [1, 7, 8]. For example, Wang et al. (2015) reported a case study, "July 2012 Beijing flood" to investigate how emergency information was timely distributed using social media during emergency events. Their research provided a better insight into the events such that decision-makers could action on emergencies promptly [7]. However, such approaches of first classifying the topics and then separately studying the topic changing with time cannot reveal how topics evolve and quantify the evolution of topics.

Herein, a novel method to reveal how topics evolve and quantify topic evolution for enhancing situation awareness in times of crisis is presented. More specifically, we propose a framework based on community evolution to analyze topic evolution over time in social media for understanding the process and stages of disasters. The framework uses communities of the co-word network in social media to represent topics. Furthermore, we propose a community evolutionary network in the framework to support and quantify the evolution of topics. Compared with existing methods, our method can reveal how topics are evolving in social media more effectively, truthfully and in detail. Our method also provides a new way to quantify and analyze the process and pattern of topic evolution in social media data. This enables researchers and disaster managers to have a deeper understanding of how the content people talk about in social

media changes with the changes of disaster situations and to effectively analyze the corresponding relationship between the topic evolution and disaster stages. Using a rainstorm that occurred in Beijing, China in July 2012 as a case study, we implemented our framework using the Sina Weibo dataset and performed a comparative analysis of the topic evolution and the stage of the actual rainfall. Our case studies show that our method can quantify the evolution process of topics and reveal the process and stages of the rainstorm. These results demonstrate that the proposed method is useful for enhancing situation awareness. The work can help disaster relief manager to quickly launch the appropriate action at the appropriate time to achieve the purpose of reducing disasters and saving lives.

This article is organized as follows: Section II describes the related studies. Section III describes the method. Section IV describes the study area and data. Section V describes the method implementation. Section VI presents the experimental results and analysis. The work and outcomes are discussed in Section VII, and summarized in Section VIII.

II. RELATED STUDIES

A. USING SOCIAL MEDIA FOR SITUATIONAL AWARENESS

Social media sites produce a large amount of user-contributed data daily. Such data have attracted the attention of many researchers who wish to apply social media data to solve intractable problems across many domains [2]. Currently, social media data has been applied to urban dynamics [9], disaster emergency [6], urban air pollution monitoring [2], election forecasting [10, 11] and so on.

Due to the openness, user orientation and real-time characteristics of social media platforms, social media has become prominent in the process of disaster management [12]. More specifically social media becomes an emergency management tool that can be used in help disaster managers disseminate right emergency alerts information to the right communities, in citizens' self-organization of local responses to a crisis, and in assessing a crisis situation (situational awareness) [13, 14], and also in facilitating responders to receive assistance requests [15].

According to Endsley (1995), situational awareness is "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [16]. Social media data, which contain useful metadata fields such as user ID, timestamp, text, and coordinates, are becoming valuable information input to support situational assessment [17].

Studies using social media for situational awareness primarily focuses on the three dimensions of social media data – time, space, and content (primarily text) [12]. Spatial information in social media data can be used in disaster mapping. Therefore, help disaster responders gain situational awareness in crisis identification and loss assessment [18, 19, 20]. The real-time characteristics of social media streams can help detect the outbreak of disasters [21, 22]. Furthermore,

researchers have used time information in social media data to understand the process of a disaster [1]. The content (primarily text) in social media data can help disaster managers to understand what had occurred during an event [12].

Typically, these three dimensions (i.e., time, space, and content (primarily text)) are inseparable when using social media data for enhancing situational awareness. Most studies regarding social media for situational awareness are based on text mining. For example, Li et al. conducted a text classifier based on a CNN to filter microblogs that are the true and real-time descriptions of rainfall, followed by visualizing the filtered microblogs as heat maps to demonstrate the effectiveness of using social media to monitor heavy rainfall events [5]. Wang et al. (2015) presented a machine learning model to classify social media messages into five topics (traffic, weather, disaster information, loss and influence, and rescue information) and performed a time-series analysis on some of these categories to reveal the process of heavy rain disasters [7].

As Huang et al. (2015) define geographic situational awareness as knowing what is happening in space [1]; we define temporal situational awareness as knowing what is happening in time. More specifically, this paper primarily focuses on revealing the process of heavy rain disasters through topic changes on social media messages over time.

B. TOPIC CHANGES IN SOCIAL MEDIA MESSAGES DURING DISASTERS

The situation of a disaster often changes over time and hence the topics of social media messages generated by social media users. Researchers indicated that the evolutions of disaster-related topics in time in social media messages can be coordinated with the process and stage of the disasters. Understanding the process and stage of the disaster can facilitate in determining appropriate rescue operations, which can improve the efficiency of disaster management.

Research methods for revealing topic evolution in document flows (social media streams) can be primarily divided into two categories. The first kind of method is based on generative probabilistic models. The representative methods include the dynamic topic model [23], topic over time [24], online topic model (OLDA) [25, 26], dual-OLDA [27], and other time series topic models [28, 29, 30, 31].

The second kind of method is based on the theory of cluster detection or called community detection in the complex system. A complex system is a system featuring many interacting components [32], which can be typically represented by network or graphs (a collection of nodes (or vertices) and connections (or edges)). A community is usually a closely connected part of a network. The detection of community structure is useful for understanding the function of the system represented by the network [33], which has applied in many fields such as online social networks, business management [34, 35, 36], communication science [37], sociology [38, 39], bibliometrics [40], and all of physical and life sciences [41].

Time as an important dimension is important issue for depicting the dynamic complex network. Several works provide how to incorporating the time dimensions in the complex network, like fusing the temporal changes while analyzing the activities in the wifi users [42] or using the power law to describe the dynamic interactions in cyber-social populations [39]. Besides, there are also researchers to investigate the dynamics of the community (community evolution) to study the dynamics of complex networks [43]. Such methods include core-based community evolution mechanism (CoCE) [44], FacetNet [45], GraphScope [46] and group evolution discovery (GED) method [47].

Community evolution of dynamic complex network can be used for the co-word network to analyze the evolution of topics [48]. The discovery of a topic (or subtopic) in social media can be achieved through the community detection of the co-word network. Furthermore, the evolution of topics can be revealed by the community evolution of the co-word dynamic network [49]. Based on the comparability identification of the co-word network community at different times, Chen and Sun (2016) divided the evolution process of subtopics into three stages: subtopic production, subtopic diffusion, and subtopic fading [49]. They discovered that representing subtopics using the co-word network community offered intelligibility and noise reduction and that the development path and variation trend of subtopics could be elucidated based on the co-word network community.

Moreover, the two methods above can be used to analyze the evolution of topics in social media during an event, thereby revealing the dynamic changes of situational awareness. Currently, most studies regarding the topic changes of social media messages under situational awareness primarily focus on the correspondence between the temporal changes of the topics and the stages of disasters. For example, Huang et al. (2015) categorized social media messages into different topics within different disaster stages and analyzed these topics trend over time. They discovered that these messages of different topics presented different volumes at different phases of the disaster [1]. Wang et al. (2015) presented a topic model of Weibo messages and investigated how different topics were timely distributed during the 2012 Beijing Rainstorm. They revealed the differences in the temporal distribution of Weibo messages of different topics before and after the rainstorm [7]. Deng et al. (2018) presented an interactive topic modeling and a streamgraph visualization method to analyze topic evolution over time to understand the dynamics of public expressions after a major explosion [8]. Fan et al. (2018) proposed a system analytics framework to detect topic evolutions associated with the performance of infrastructure systems for tracking the movement of situations in different disaster phases [50]. Using the Central European Flooding 2013 as a case study, Grunder-Fahrer optimized and applied topic model analysis and temporal clustering techniques to investigate the thematic and temporal structure of German social media communications [51].

From the above mentioned studies, it is clear that dynamic

situational awareness can be obtained by tracking the evolution of topics in social media data during a disaster. Studies of topic evolution analysis in computer science primarily track topic changes by mining the topic life cycle and its evolution model. However, studies in disaster management are primarily based on the temporal changes of social media topics. Few studies focus on using the evolution model to reveal the change in topics and the corresponding relationship with the process and stage of disasters.

III. METHODOLOGY

We propose a framework to analyze the topic evolution in social media for enhancing situation awareness in times of crisis. The framework uses communities of the co-word network in social media to represent topics. Based on the evolution of the communities, we propose a community evolutionary network to support the quantification of topic evolution. As shown in Fig. 1, the framework involves the following steps.

- 1) The time interval during a disaster is divided into successive time segments by the time window (see Fig. 1 (a)).
- 2) In each time window, a co-word network is created on social media messages and then the community of each co-word network is detected (see Fig. 1 (b)).
- 3) The evolutionary events between communities in adjacent periods are identified (see Fig. 1 (c)).
- 4) The community evolutionary network is constructed according to the community evolution. We can apply the method of network analysis to the community evolutionary network to quantify the evolution of topics (see Fig. 1 (d)).

A. TIME WINDOWS

Time windows T is a series of successive connected time intervals of equal length, which is defined as (1) and (2):

$$T = \{T_1, T_2, T_3, \dots, T_n\} \quad (1)$$

$$T_{length} = \|T_1\| = \|T_2\| = \|T_3\| = \dots = \|T_n\|, \quad (2)$$

where $T_1, T_2, T_3, \dots, T_n$ are different time windows, T_{length} is the length of each time window.

B. CO-WORD NETWORK AND COMMUNITY DETECTION

1) Co-word network

Within each time window, we build a co-word network (see Fig. 1 (b)). The co-word network uses topic words as nodes. For each of two topic words, if they appear in the same social media messages, then an edge is created between these two words and the weight of the edge is the frequency of the two words that appear in all documents.

To build a co-word network, we first identify the topic words based on the term frequency-inverse document frequency (TF-IDF), which shows the importance of a word to

a document [52]. The TF-IDF of a word t_i in document d_j (denoted as $tfidf_{i,j}$) can be calculated according to (3):

$$tfidf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|D|}{|\{j : t_i \in d_j\}|}, \quad (3)$$

where $n_{i,j}$ is the number of occurrences of a word t_i in document d_j (one Weibo in our case), $\sum_k n_{k,j}$ is the sum of occurrences of all words in a document d_j , $|D|$ represents the total number of documents (Weibos) in the dataset, and $|\{j : t_i \in d_j\}|$ means the number of documents that contain the word t_i .

We calculate the importance of a word t_i to all documents (i.e. all Weibos) $\{D\}$ (denoted as W_i) as the sum of TF-IDF values of a word in all documents using (4); subsequently, the topic words can be identified using (5).

$$W_i = \sum_j tfidf_{i,j} \quad (4)$$

$$\{t_{topic}\} = \{t_i \mid W_i \geq \theta\}, \quad (5)$$

where $\{t_{topic}\}$ is the topic words to be identified, and θ the threshold that must be set.

2) Community detection

We applied the Louvain method for community detection (LM) algorithm [53] to conduct community detection on the constructed co-word network. Once the community is detected, the topic can be identified by examining the topic words in the community.

The LM algorithm is a modularity-based method. Modularity measures the strength of the division of a network into communities. The modularity-based community detection algorithm is aimed at maximizing the modularity. The LM algorithm optimizes the modularity in two steps that are repeated iteratively. First, small communities are obtained by optimizing the modularity locally on all nodes in the network; subsequently, each small community is grouped into one node (which can be regarded as a new network), and the first step is repeated. This process is applied repeatedly until the modularity stops increasing.

The LM algorithm offers several advantages that justify its use for detecting the community of the co-word network. The algorithm is extremely fast and its complexity is linear on typical and sparse data. In other words, it can be used for extracting communities from large networks. Furthermore, the algorithm circumvents the resolution limit problem [54] of modularity owing to its intrinsic multilevel nature.

C. TOPIC EVOLUTION IDENTIFICATION

The topic evolution was derived from the community evolution of the co-word network. Community evolution (or group evolution) is a sequence of events (changes) of communities succeeding each other in consecutive time windows within the temporal network. Fig. 1 (c) shows three possible evolutionary events from T_1 to T_2 . The number, name, and definition

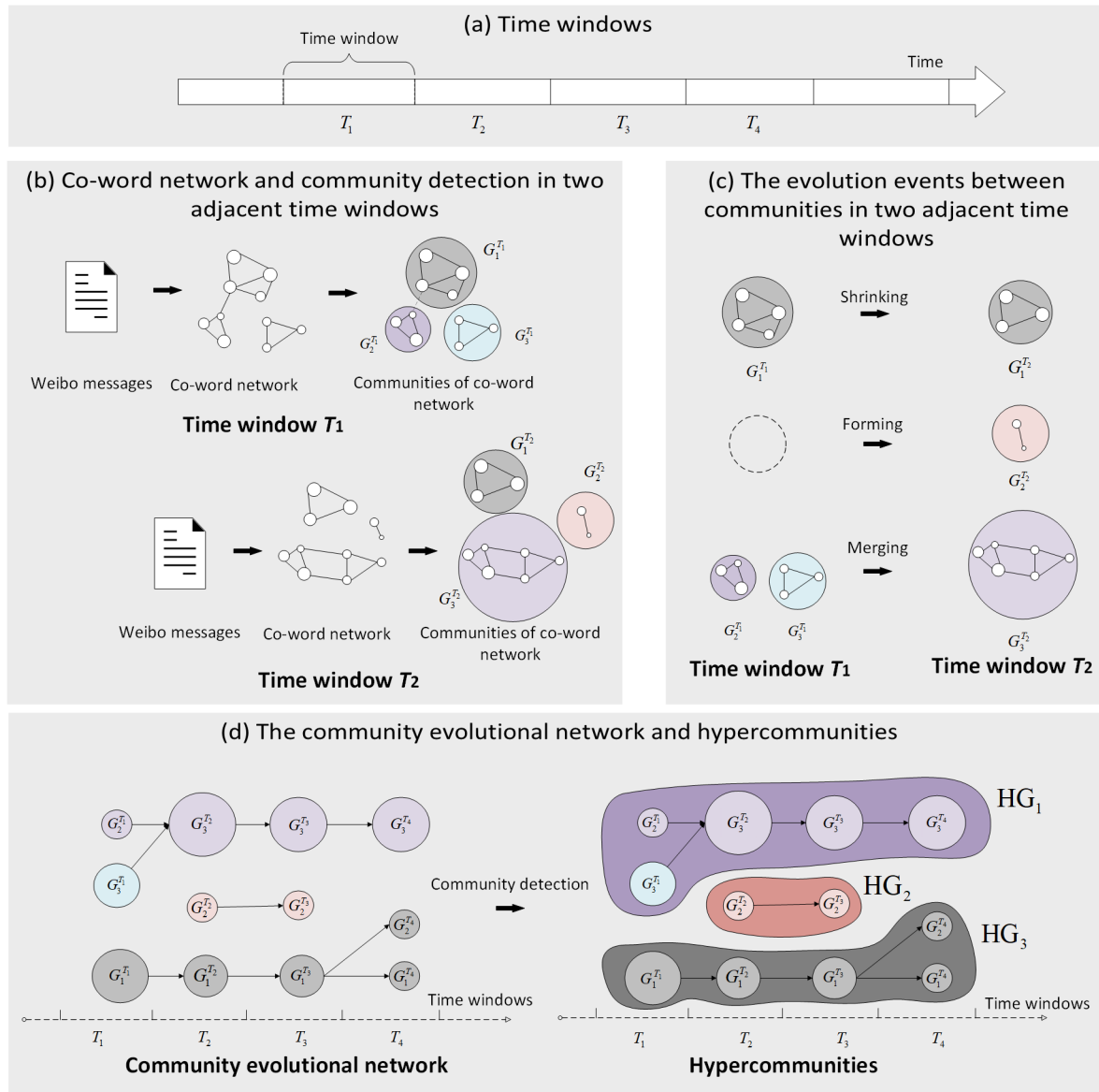


FIGURE 1: Schematic of the topic evolution analysis framework. (a) Time windows. One time interval during a disaster is divided into 4 successive time window T_1, T_2, T_3, T_4 . (b) Co-word network and community detection in two adjacent time windows T_1 and T_2 . In each time window T_1 and T_2 , a co-word network is created on social media messages and then the community of each co-word network is detected. $G_1^{T_1}, G_2^{T_1}$ and $G_3^{T_1}$ are communities in time windows T_1 , $G_1^{T_2}, G_2^{T_2}$ and $G_3^{T_2}$ are communities in time windows T_2 . (c) The evolution events between communities in two adjacent time windows T_1 and T_2 . Community $G_1^{T_1}$ is shrinking to $G_1^{T_2}$ from T_1 to T_2 , G_2 is forming in T_2 , $G_2^{T_1}$ and $G_3^{T_1}$ are merging into $G_3^{T_2}$ from T_1 to T_2 . (d) The community evolutionary network and hypercommunities. The node $G_i^{T_j}$ of the community evolutionary network is the community G_i of the co-word network in time windows T_j . The edge of community evolutionary network is the evolution events between communities in two adjacent time windows. HG_i indicates the hypercommunity, which is the community of community evolutionary network.

of the events differ according to studies; however, all of them are highly similar and complete each other [43]. We adopted seven types of events to describe the changing state of a community or communities between two consecutive times from [47]. These events include continuing, shrinking, growing, splitting, merging, dissolving, and forming. The definition of these events can be found in Table 1.

We adopted a simplified version of the group evolution discovery (GED) method [47] to identify the category of community evolution events. The method uses a measure called inclusion to evaluate the inclusion of one community in another. The inclusion $I(G_1, G_2)$ of community G_1 in community G_2 is calculated using (6):

$$I(G_1, G_2) = \frac{|G_1 \cap G_2|}{|G_1|} \cdot \frac{\sum_{t \in (G_1 \cap G_2)} WD_{G_1}(t)}{\sum_{t \in (G_1)} WD_{G_1}(t)} \quad (6)$$

$|G_1|$ is the quantity of topic words in community G_1 , $|G_1 \cap G_2|$ the quantity of the topic words that are shared by community G_1 and G_2 , and $WD_{G_1}(t)$ the weighted degree of node t in community G_1 .

Once the inclusions $I(G_1, G_2)$ and $I(G_2, G_1)$ are calculated, the community evolution events between G_1 and G_2 can be identified according to Table 1.

Here, α and β are two thresholds that must be set. $M_{G_1}^{T_{i+1}}$ is the number of matches between G_1 and all groups in the next time window T_{i+1} . It is noteworthy that dissolving and forming events are actual events to describe community G_1 and G_2 separately. The events dissolving for G_1 and forming for G_2 can appear simultaneously.

D. TOPIC EVOLUTION QUANTIFY BASED ON COMMUNITY EVOLUTIONAL NETWORK

We propose a community evolutionary network for quantifying the evolution of the community, which can help to quantify the changing process of topics in social media. The community evolutionary network is defined as (7):

$$CE - Network = \{HN^T, HE, Weight\}, \quad (7)$$

where HN is the node of the community evolutionary network, defined as the hypernode for distinguishing with the node of the co-word network. The hypernodes are the communities detected in co-word networks. T is the time window that hypernode HN belongs to, and HN^T is a hypernode in time window T . HE is the directed edge of the community evolutionary network, defined as the hyperedge for distinguishing with the edge of the co-word network. If any evolution events occur except dissolving and forming between two communities in the co-word network, a hyperedge between their corresponding hypernodes is created. $Weight$ is the weight of the hyperedge. The weight of a hyperedge $HE(HN_1^{T_i}, HN_2^{T_{i+1}})$ is the number of nodes in two communities corresponding to two hypernodes connected by the hyperedge, which can be calculated using (8).

$$Weight_{HE(HN_1^{T_i}, HN_2^{T_{i+1}})} = |G_1^{T_i} \cap G_2^{T_{i+1}}|, \quad (8)$$

where $G_1^{T_i}$ and $G_2^{T_{i+1}}$ are the corresponding communities in the co-word network of hypernodes $HN_1^{T_i}$ and $HN_2^{T_{i+1}}$, respectively.

The community evolutionary network is a temporal direct network; therefore, the methods for direct network analysis can be applied to the community evolutionary network. We present two components to quantitatively analyze community evolution using the community evolutionary network.

- 1) Hypercommunity HG . The communities of the community evolutionary network are called hypercommunities. The detection of hypercommunities can be the same as that of the co-word network (i.e., the LM algorithm).
- 2) Lifetime τ and life-span $|\tau|$. The lifetime of a hypercommunity HG is the set of time windows of all hypernodes in this hypercommunity, which can be defined using (9):

$$\tau_{HG} = \{T_m, T_{m+1}, \dots, T_n\}, n > m \quad (9)$$

The life-span $|\tau_{HG}|$ of a hypercommunity HG is the duration of this hyper-community, which can be calculated using (10):

$$|\tau_{HG}| = |m - n| \cdot |T_{length}| \quad (10)$$

IV. CASE STUDY AND DATA

A. CASE STUDY

We selected “July 2012 Beijing flood” as our case study. From July 21st to 22nd, 2012, heavy rain occurred in many parts of China. Among them, Beijing and its surrounding areas experienced the strongest rainstorm in 61 years. The heavy rain caused large areas of flood in Beijing [55]. As of August 6, 79 people have died in the rainstorm in Beijing. According to data from the Beijing Municipal Government, the storm caused 1.602 million people to be affected, 10,660 houses to collapse, and an economic loss of 11.64 billion yuan [55].

The torrential rain involved two stages [56]. The first stage was from 10 a.m. to 8 p.m. (LTC) on 21 July 2012, which presented convective precipitation with short-time, high rainfall intensity and obvious fluctuations. The second stage was from 8 p.m. on 21 July to 4 a.m. on 22 July 2012, which presented smooth frontal precipitation.

B. DATA

Sina Weibo data were collected by combining a web crawler with the Sina Weibo application programming interface, which refers to the data collection method used in [2]. We collected a sample of Weibos using the keywords “北京暴雨” (Beijing Rainstorm), yielding 389,168 Weibos for Beijing 6:00 p.m. on July 21 to 4 p.m. on July 24. Only geotagged Weibos were used in this study to filter retweeted messages and noises such as Weibos used for marketing and

TABLE 1: Identifying conditions of community evolution events. $I(G_1, G_2)$ indicates the inclusion of community G_1 in G_2 , $I(G_2, G_1)$ indicates the inclusion of community G_2 in G_1 . $|G_1|$ is the number of topic words in community G_1 , $|G_2|$ is the number of topic words in community G_2 . $M_{G_1}^{T_{i+1}}$ is the number of matches between G_1 and all groups in the next time window T_{i+1} , $M_{G_2}^{T_i}$ is the number of matches between G_2 and all groups in the previous time windows T_i . α and β are two thresholds that must be set.

Community evolution Events	Definition	Identifying Conditions
Continuing	no significant differences in the consecutive time windows	$I(G_1, G_2) \geq \alpha$ and $I(G_2, G_1) \geq \beta$ and $ G_1 = G_2 $
Shrinking	some nodes left the community	$I(G_1, G_2) \geq \alpha$ and $I(G_2, G_1) \geq \beta$ and $ G_1 > G_2 $ OR $I(G_1, G_2) < \alpha$ and $I(G_2, G_1) \geq \beta$ and $ G_1 \geq G_2 $ and $M_{G_2}^{T_i} = 1$
Growing	some new nodes joined the community	$I(G_1, G_2) \geq \alpha$ and $I(G_2, G_1) \geq \beta$ and $ G_1 < G_2 $ OR $I(G_1, G_2) \geq \alpha$ and $I(G_2, G_1) < \beta$ and $ G_1 \leq G_2 $ and $M_{G_1}^{T_{i+1}} = 1$
Splitting	a community splits into several communities	$I(G_1, G_2) < \alpha$ and $I(G_2, G_1) \geq \beta$ and $ G_1 \geq G_2 $ and $M_{G_2}^{T_i} > 1$
Merging	a community was created by merging several other communities	$I(G_1, G_2) \geq \alpha$ and $I(G_2, G_1) < \beta$ and $ G_1 \leq G_2 $ and $M_{G_1}^{T_{i+1}} > 1$
Dissolving	a community does not occur in the next time window	for G_1 in T_i and each group G_2 in T_{i+1} , $I(G_1, G_2) < \alpha$ and $I(G_2, G_1) < \beta$
Forming	a new community (which has not existed in the previous time window) occurs	for G_2 in T_{i+1} and each group G_1 in T_i , $I(G_1, G_2) < \alpha$ and $I(G_2, G_1) < \beta$

advertising. A total of 16,759 geotagged Weibos remained. Each Weibo record contains information such as text message, post time, and geographic coordinates.

V. METHOD IMPLEMENTATION AND COMPUTATIONAL COMPLEXITY ANALYSIS

A. PARAMETER SETTINGS

According to the description presented in the method section, to implement our method on the Weibo dataset, we must set the threshold θ to identify the topic words; set the length of the time window T_{length} ; and set the threshold of the inclusion, α and β , to detect the topic evolution events.

1) The threshold θ and topic words

We identified the threshold θ of W by the head/tail break method [57]. For data with power-law distribution, it can be classified by the head/tail break method. Baayen (2001) reported that the distribution of word frequencies in a document primarily conformed to the heavy-tailed distribution, which is a typical distribution in nature [58]. For our data, we observed whether the distribution of W obeyed the power law; if it does, then we use the head/tail break method to classify W and find the boundaries of the classification to determine the threshold θ of W and the topic words.

Most Weibo messages are written in Chinese. Because the Chinese language does not involve a trivial word segmentation process, we must first apply text segmentation to calculate W . This study was implemented using the Jieba¹ Python Chinese word segmentation module. Typically used meaningless words such as “的”, “了” are used as stop words during text segmentation. Once text segmentation is completed, the value of W_i can be calculated according to (3) and (4). The distribution of W is shown in Fig. 2.

¹Chinese text segmentation tool, <https://github.com/fxsjy/jieba>

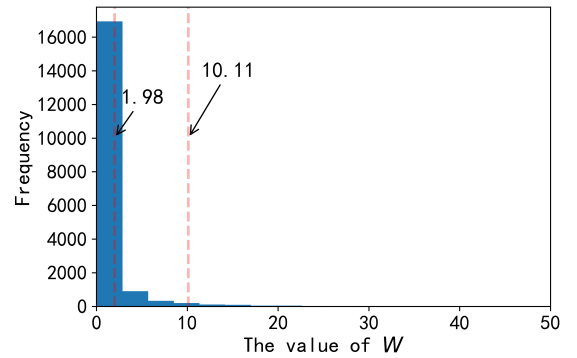


FIGURE 2: Distribution of the value of W . The x-axis indicates the value of W , the y-axis the frequency of the value of W . The red vertical lines indicate the two thresholds θ obtained by the head/tail break method.

It is clear that the W value of most characters in Weibo messages is relatively small, and only the W value of a small part of the characters is relatively large. The distribution of W obeys the power-law distribution. We used the head/tail break method to perform two head-to-tail breaks on the data and discovered that the head no longer obeyed the power-law distribution. The words of the last head (i.e., $\theta \geq 10.11$) were selected as the topic words, and a total of 654 topic words were obtained.

2) The length of time windows T_{length}

We set the length of each time window as 2 hours. [7] explored the changes of disaster topics in Weibo messages using the time window length of 1 hour. Referring to [7], we set 2 hours as our time window length. In this case, the Weibos in each time window are sufficient to guarantee

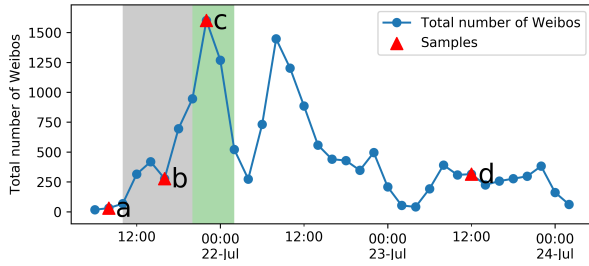


FIGURE 3: Total Weibos in each time window. The x-axis indicates date and time, the y-axis the total Weibos in each time window. The gray vertical span (rectangle) indicates the first stage of torrential rain, the green vertical span (rectangle) the second stage of torrential rain. The red marker (denoting a, b, c, and d) indicates randomly selected samples of the time windows before, in the first stage, in the second stage, and after the torrential rain, respectively.

statistically significant analysis results under a narrower time window. The total Weibos in each time window is shown in Fig. 3.

3) The thresholds of inclusion α and β

The distribution of inclusion I is shown in Fig. 4. It is clear that the distribution of inclusion I obeys the power law. We used the head/tail break method [57] to set the threshold. After two head-to-tail breaks on the data, when α less was than 0.0383, and β less than 0.039, the head no longer obeyed the power-law distribution. Therefore, we set α, β as 0.0383, 0.039.

B. COMPUTATIONAL COMPLEXITY ANALYSIS

The proposed method is mainly implemented using python and python-related packages, most of which are widely used in data analysis and mining. The packages and software involved in the method implementation, the corresponding code running time, and running environment, the computational complexity of each step of the proposed method are shown in Table 2.

The proposed method took a total of 28.109 seconds, which have an acceptable run time. The primary run time of the method is spent on step 2 (24.1 seconds) and step 4 (2.61 seconds).

The computational complexity of step 2 (the creation of the co-word network in all time windows) is $O(T \cdot nL^2)$, where T is the total number of time windows, n the average number of documents (Weibo messages) in all time windows, and L the average length of documents (Weibo messages). Since Sina Weibo limited the number of characters in each posted message to 140 before January 2016 [63], the average length of documents (Weibo messages) is less than 140. The total number of time windows is related to the application situation and the length setting of time windows. Our application takes 2 hours as the length of the time window, and gets 35

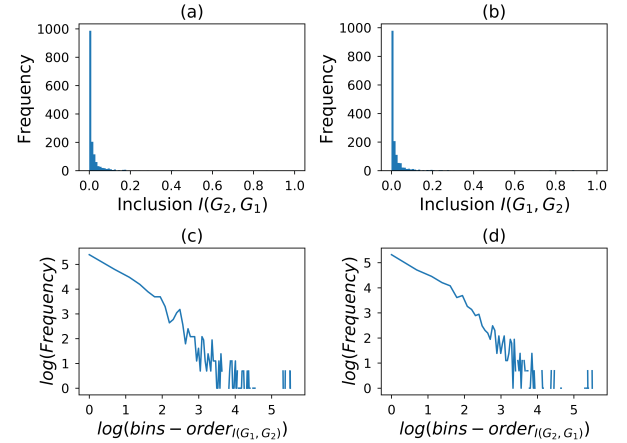


FIGURE 4: Distribution of the inclusion I . The x-axis of Fig. (a) and (b) indicate the value of inclusion I , the y-axis of Fig. (a) and (b) indicate frequencies of the value of inclusion I . The x-axis of Fig. (c) indicates the natural logarithm of the bins order in Fig. (a), the y-axis of Fig. (c) indicates the natural logarithm of the frequencies of the inclusion I . The x-axis of Fig. (d) indicates the natural logarithm of the bin order in Fig. (b), the y-axis of Fig. (d) the natural logarithm of the frequencies of the inclusion I .

time windows. The total number of time windows in most applications that use social media for disaster situational awareness is usually not too large. According to the above analysis, the run time of step 2 (the creation of the co-word network in all time windows) mainly depends on the total number of documents (Weibo messages) in the dataset.

The computational complexity of step 4 (community evolution events identification) is $O(T \cdot n^2)$, where T is the total number of time windows, n is the average number of communities in all time windows. Because T is usually not a large value, the run time of step 4 mainly depends on the average number of communities in all time windows. In the case of co-word network, the number of communities is also usually not a large value. Therefore, step 4 (community evolution events identification) is not difficult to execute in most cases of the co-word networks.

VI. RESULTS

A. TOPIC DISCOVERY

To verify the validity of the time window setting and topic discovery method, we randomly selected four time windows at each stage of the rainstorm (see Fig. 3 and 5) and analyzed the topics of the communities detected in each time window. Fig. 5 shows the results of the communities in four selected time windows.

We discovered that most of the communities can be mapped to disaster-related topics. The topics of each community were manually identified by identifying the topic words and Weibo messages related to each community (Table 3).

Furthermore, we discovered that the topics differ at dif-

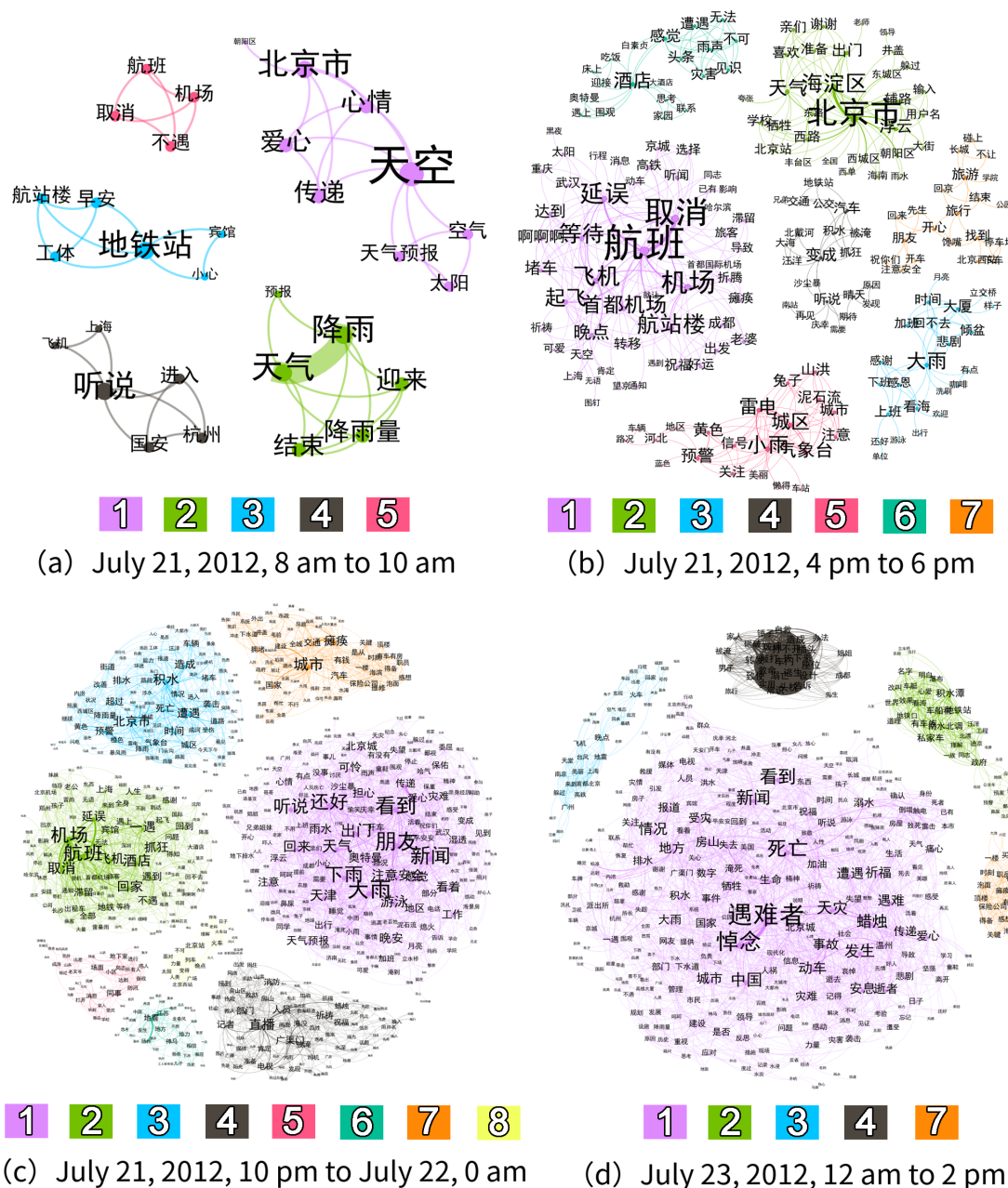


FIGURE 5: The detected communities of co-word networks in four randomly selected time windows. The size of the node (topic words) indicates its degree. The main Chinese topic words in each community were translated into English, which are shown in Table 3.

TABLE 2: The computational complex, implementation language and packages, and CPU running times of the proposed method's main steps. The running environment of the implementation is — CPU model: Intel(R) Xeon(R) CPU E3-1240 v5 @ 3.50GHz; CPU kernels: 8 (the implementation runs on 1 kernel); CPU operating modes: 32-bit, 64-bit; System: Ubuntu 16.04.1 64-bit; Software: Python 3.6.9 :: Anaconda, Inc (<https://www.anaconda.com>).

Main steps of the implementation	Computational time complexity	Implementation language and packages	CPU running times (seconds)
1.The calculation of $tfidf$ and W	$O(nL \log nL)$, where n is the total number of documents (Weibo messages) in the dataset and L the average length of documents (Weibo messages)	Python, Scikit-learn [59]	0.915
2.The creation of the co-word network in all time windows	$O(T \cdot nL^2)$, where T is the total number of time windows, n the average number of documents (Weibo messages) in all time windows, and L the average length of documents (Weibo messages)	Python, Numpy [60], Networkx [61]	24.1
3.Community detection of co-word networks in all time windows	$O(T \cdot n \log n)$, where T is the total number of time windows, n is the average number of nodes of networks in all windows	Python, Networkit [62]	0.15
4.Community evolution events identification	$O(T \cdot n^2)$, where T is the total number of time windows, n is the average number of communities in all time windows	Python	2.61
5.The creation of community evolutionary network	$O(n)$, where n is the total number of community evolution events	Python, Networkx [61]	0.228
6.Supercommunities detection	$O(n \log n)$, where n is the total number of nodes of the community evolutionary network	Python, Networkit [62]	0.106
In total			28.109

ferent time windows, which can reflect the stages of the rainstorm. As shown in Fig. 5 (a) and Table 3, only a few topic words appeared before the rainstorm. Nonetheless, we could still obtain topics regarding the weather forecast that is related to the rainstorm (communities 1 and 2) and the effect of this weather forecast on flights (community 3, 4, and 5).

More topic words appear in the first stage of the rainstorm (see Fig. 5 (b) and Table 3). Owing to torrential rain, the community of the topic on flight delays and cancellation became the largest community (Fig. 5 (b) Community 1). In addition, a large number of Weibo users began to post water-related and traffic-related messages on Weibo (Fig. 5 (b) Community 4). Furthermore, we discovered that traffic problems affected the normal commute (Fig. 5 (b) Community 3). Additionally, traffic problems affected tourists. Therefore, many keywords about tourism and hotels appeared in Weibo messages (Fig. 5 (b) Community 6 and Community 7). Furthermore, owing to torrential rain, the government began to publish information on flash floods and mudslide warnings on Weibo (Fig. 5 (b) Community 5). It is clear that in a rainstorm, people will post messages related to the rainstorm location (Fig. 5 (b) Community 2).

In the second phase of the rainstorm (see Fig. 5 (c)), more keywords appeared in Weibo than in the first phase of the storm. The topic of the largest community (Fig. 5 (c) in community 1) is friends. As the rainstorm situation becomes more urgent, news related to the rainstorm begins to increase, and people begin to focus more on their friends in Beijing (Fig. 5 (c) in community 1). In addition, traffic-related topics continue. For example, the topic of flight cancellation and delay (Fig. 5 (c) community 2) continued at this stage, and the topics related to train delays begin to appear; furthermore, people begin to discuss the reasons for traffic congestions during the rainstorm (Fig. 5 (c) community 7). It is clear that information regarding stagnant water and casualties begin to appear in a community, thereby forming a new topic (Fig.

5 (c) community 3). This reflects people's casualties caused by water accumulation. Furthermore, rescuers begin implementing rescue activities at this stage (Communities 4 and 5 in Fig. 5 (c)). Finally, news regarding the Jiangsu Yangzhou earthquake at 20 o'clock on July 20, 2012 is discussed in Weibo [64], which reflects the fact that topics discussed in Weibo are affected by adjacent time events.

After the rainstorm, the topics of the Weibo messages primarily contain four parts. 1) Mourning and blessing for victims, such as in the largest community (community 1) in Fig. 5 (d); 2) discussion of the disaster (Fig. 5 (d) community 2 and 7); 3) traffic problems such as delayed trains and airplanes (Fig. 5 (d) community 3); 4) Discussion of self-seeking methods (Fig. 5 (d) community 4).

B. TOPIC EVOLUTION ANALYSIS

Fig. 6 shows the distribution of different evolution events over time. In general, the majority of the evolution events involve forming and dissolving appearing simultaneously, followed by merging and splitting events. The remaining events such as continuing, growing, and shrinking are relatively few. This reflects the diversity of topics and the rapid evolution of the topic during the rainstorm. Furthermore, we discovered that the trend of the number of topic evolution events over time reflected the stage and process of the rainstorm. The forming and dissolving events began to increase as the rainstorm started and decreased as the storm ended (Fig. 6). By examining Weibo messages, we discovered that people posted messages related to the rainstorm in Weibo primarily because they were affected by heavy rain. In addition, after the storm ended, the forming and dissolving events began to increase and continued for a while and then slowly decreased (Fig. 6). By examining Weibo messages, we discovered that people discussed this rainstorm in Weibo primarily because they were affected by the news they heard.

TABLE 3: Main topic words (both Chinese and their English translations) and corresponding topics of co-word network communities in Fig. 5.

Communities	Topic words	Corresponding topic
Fig. 5 (a) community 1	天空(sky), 天气预报(weather forecast), 太阳(sun)	weather forecast
Fig. 5 (a) community 2	天气(weather), 降雨(rainfall), 降雨量(precipitation)	weather forecast
Fig. 5 (a) community 3	地铁站(subway station), 航站楼(terminal building)	flight
Fig. 5 (a) community 4	听说(I heard), 进入(enter), 飞机(plane)	flight
Fig. 5 (a) community 5	航班(flight), 取消(cancellation), 机场(airport)	flight
Fig. 5 (b) community 1	航班(flight), 取消(cancellation), 延误(delay)	flight delays and cancellation
Fig. 5 (b) community 2	北京市(Beijing), 海淀区(Haidian District), 西路(West road)	locations
Fig. 5 (b) community 3	大雨(heavy rain), 加班(working overtime), 上班(go to work), 下班(get off work)	commuting
Fig. 5 (b) community 4	积水(water), 被淹(flooded), 汽车(cars), 变成(turned into)	water and traffic
Fig. 5 (b) community 5	城区(urban), 气象台(meteorological observatory), 泥石流(debris flow), 预警(warning)	mountain torrents, mudslides
Fig. 5 (b) community 6	酒店(hotel), 遭遇(encounter, suffer), 灾害(disaster)	hotel
Fig. 5 (b) community 7	旅行(travel), 旅游(tourism), 朋友(friends), 开心(happy)	travel
Fig. 5 (c) community 1	看到(I saw), 新闻(news), 朋友(friends), 听说(I heard), 注意安全(pay attention to safety)	friends
Fig. 5 (c) community 2	机场(airport), 航班(flight), 取消(cancellation), 延误(delay)	flight delays and cancellation
Fig. 5 (c) community 3	积水(water), 造成(cause), 死亡(death), 遭遇(encounter, suffer)	water and death
Fig. 5 (c) community 4	直播(live), 记者(reporter), 救助(rescue), 祈祷(pray)	live, rescue, pray
Fig. 5 (c) community 5	同事(colleague), 地下室(basement), care for friend	rescue, care for friends
Fig. 5 (c) community 6	地震(earthquake), 江苏(Jiangsu), 扬州(Yangzhou)	2012 Yangzhou earthquake
Fig. 5 (c) community 7	城市(city), 交通(traffic), 瘫痪(paralysis), 国家(country)	discussion on the urban and transform system
Fig. 5 (c) community 8	北京站(Beijing Railway Station), 火车(train), 晚点(late, delay)	train delay
Fig. 5 (d) community 1	遇难者(victims), 死亡(death), 悼念(mourning)	mourning and blessing
Fig. 5 (d) community 2	积水潭(Jishuitan), 车船税(vehicle and vessel tax), 政府(government), 南水北调(South-to-North water transfer)	discussion regarding the disaster
Fig. 5 (d) community 3	火车(trains) 晚点(late) 飞机(airplane)	traffic problem
Fig. 5 (d) community 4	自救(self-help), 逃生(escape), 设计(design), 转发(re-post)	discussion regarding self-seeking method
Fig. 5 (d) community 7	保险公司(insurance company), 瘫痪(paralysis), 顶楼(top floor), 一楼(first floor), 感想(feelings)	discussion regarding the disaster

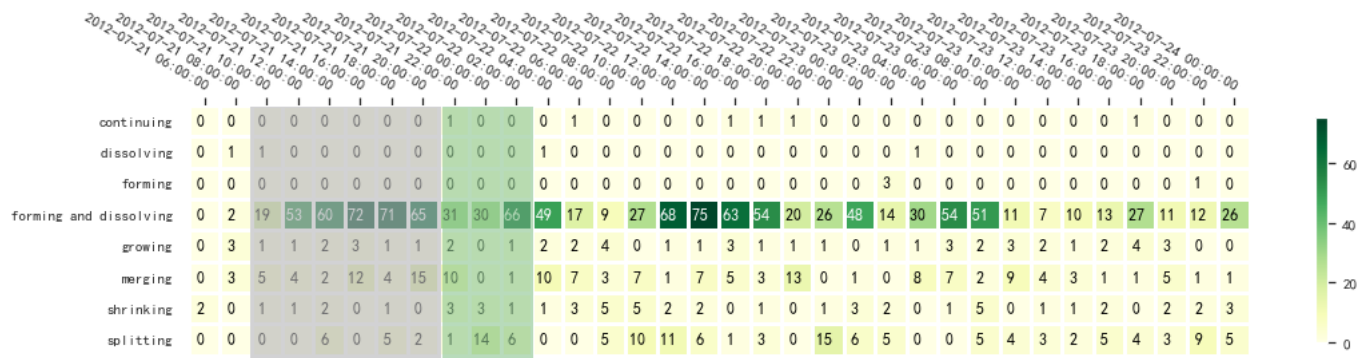


FIGURE 6: The total quantity distribution of different topic evolution events. Forming and dissolving represent the forming and dissolving events that appear at the same pair of communities, respectively. The gray vertical span (rectangle) indicates the first stage of torrential rain, the green vertical span (rectangle) the second stage of torrential rain.

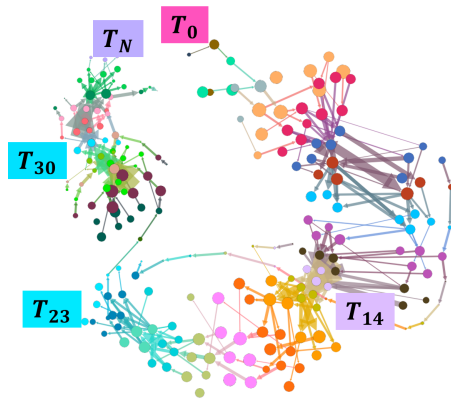


FIGURE 7: The community evolutionary network built on the rainstorm Weibo data. The different colors represent different time windows. Starting with T_0 , the community begins to evolve and ends at T_{34} .

C. TOPIC EVOLUTION QUANTIFYING

1) Community evolutionary network

To further analyze the evolution of the topic, we constructed a community evolutionary network and analyzed the communities (hypercommunities) in the community evolutionary network. The constructed community evolutionary network is shown as Fig. 7, which contains 237 hypernodes and 326 hyperedges. As shown in Fig. 7, the evolution of the topics is phased. For example, from T_0 to T_{14} (6 pm on July 21 to 8 am on July 22), the connections between the communities in this time interval are relatively close. This time period is consistent (corresponding) with the time period when the rainstorm occurs (The rainstorm started from 10 pm on July 21 to 2 am on July 22). In addition, from T_{14} to T_{23} (8 am on July 22 to the 0 am of July 23), T_{23} to T_{30} (0 am on July 23 to 4 pm on July 23) and T_{30} to T_{34} (4 pm on July 23 to 4 pm on July 24), the connections between the communities in these time intervals are relatively close as well.

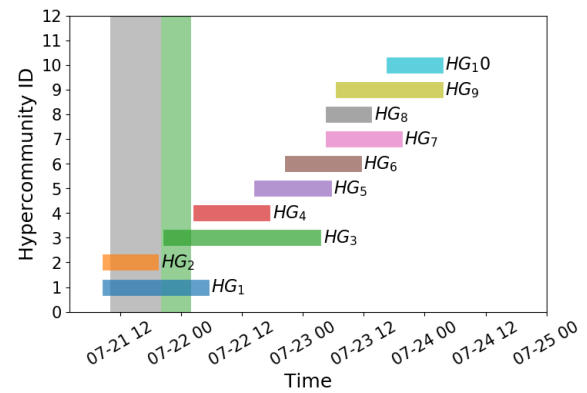


FIGURE 8: Life cycle of the hypercommunities. The gray vertical span (rectangle) indicates the first stage of torrential rain, the green vertical span (rectangle) the second stage of torrential rain.

2) Hypercommunities and lifetime

The community evolutionary network can support the quantitative analysis of topic evolution by the analysis of the communities (hypercommunities) in the community evolutionary network. Fig. 9 shows the community of the community evolutionary network (hypercommunity). Fig. 8 shows the lifetime of these hypercommunities.

As shown in Fig. 8, different hypercommunities appeared in different rainstorm phases. For example, the lifetime of hypercommunity 1 coincides with the rainfall process of the rainstorm. The lifetime of hypercommunity 2 coincides with the first stage of the rainstorm.

Additionally, we discovered certain evolution patterns of topics in different hypercommunities. According to the topology of these hypercommunities (see Fig. 9), we divided the evolution of the topics in these hypercommunities into three categories: Hypercommunities 3, 8, and 9 are line-like network without branches, i.e., they represent line-like evolutions; hypercommunities 2 and 6 are tree-like networks with branches but no ring, i.e., they represent tree-like evolutions;

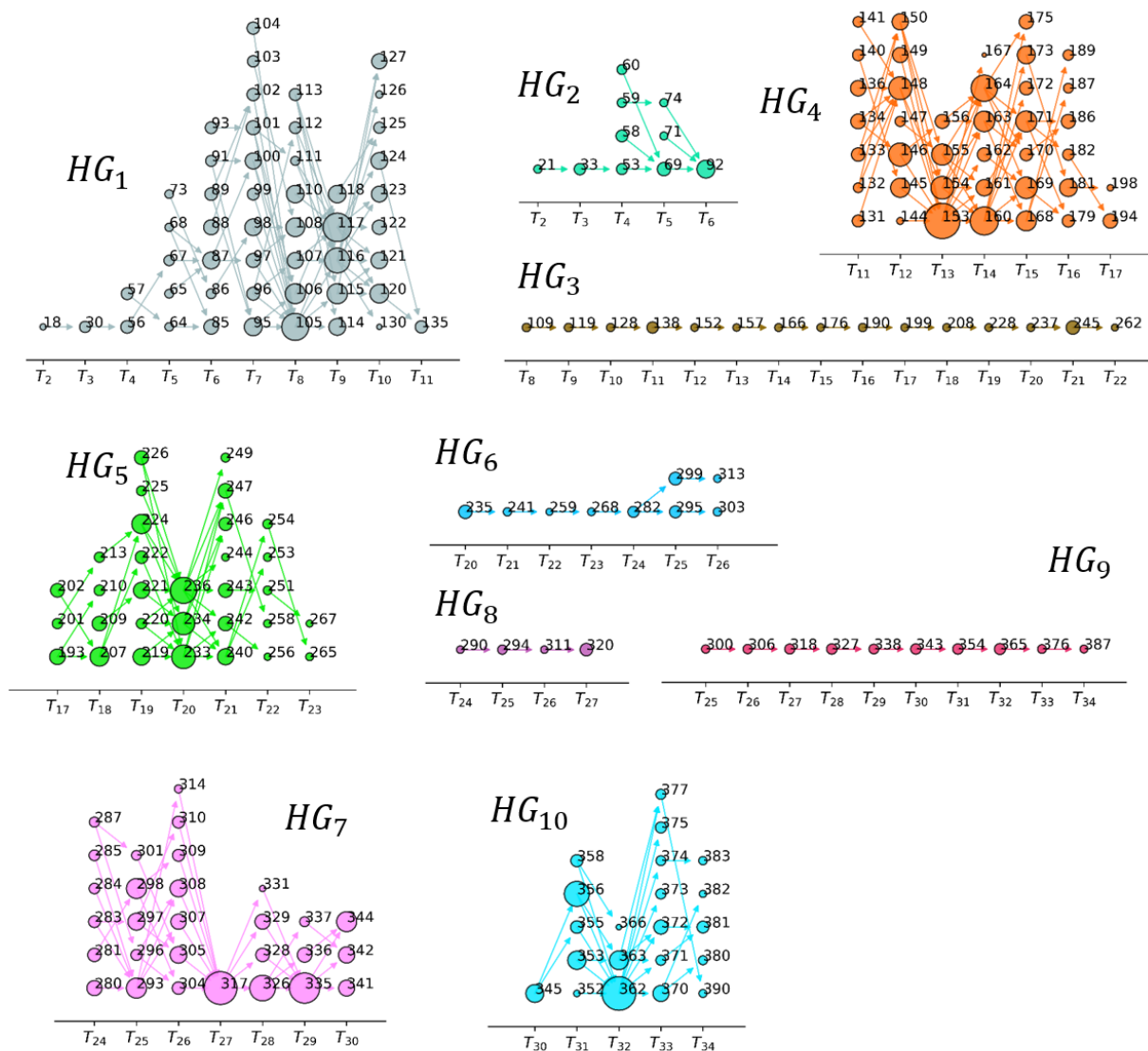


FIGURE 9: Hypercommunities of community evolutionary network. The x-axis indicates time windows T . HG_i indicates Hypercommunities i . The time windows start from T_0 (6 pm on July 21) to T_{34} (4 pm on July 24) with two hours time windows length. Each node of the hypercommunities is labeled by a unique number.

hypercommunities 1,4,5,7, and 10 are networks with rings, i.e., they represent network-like evolutions.

To further understand the meaning of different patterns of and what had happened on these hypercommunities, we examined the topics corresponding to each node in the hypercommunity. In a hypercommunity, each node is a co-word network (a community of a co-word network). Fig. 10 shows four examples of the co-word networks corresponding to these hypercommunities, and the topics of all the hypercommunities are described in Table 4.

As shown in Fig. 10 and Table 4, the line-like hypercommunity is typically formed by Weibo users who wish to spread a specific message. For example, in hypercommunity 9 (see Fig. 10 (d)), because of the incident where a driver

was trapped and drowned owing to water in a road area, many Weibo users began to spread information regarding the method to escape similar situations. Because users are basically copying and pasting when they spread such messages, the resulting community of the co-word network is close and forms a line-shaped evolution.

The tree-like communities reflect the evolution of popular topics related to disasters. For example, hypercommunity 6 was formed by the news reports of casualties. As shown in Fig. 10 (c), after T_{24} , the topic splits into two topics and then slowly disappears.

The network-like communities typically reflect the diversity of the topics. For example, hypercommunity 1 is formed by various topics. As shown in Fig. 10 (a), the topics

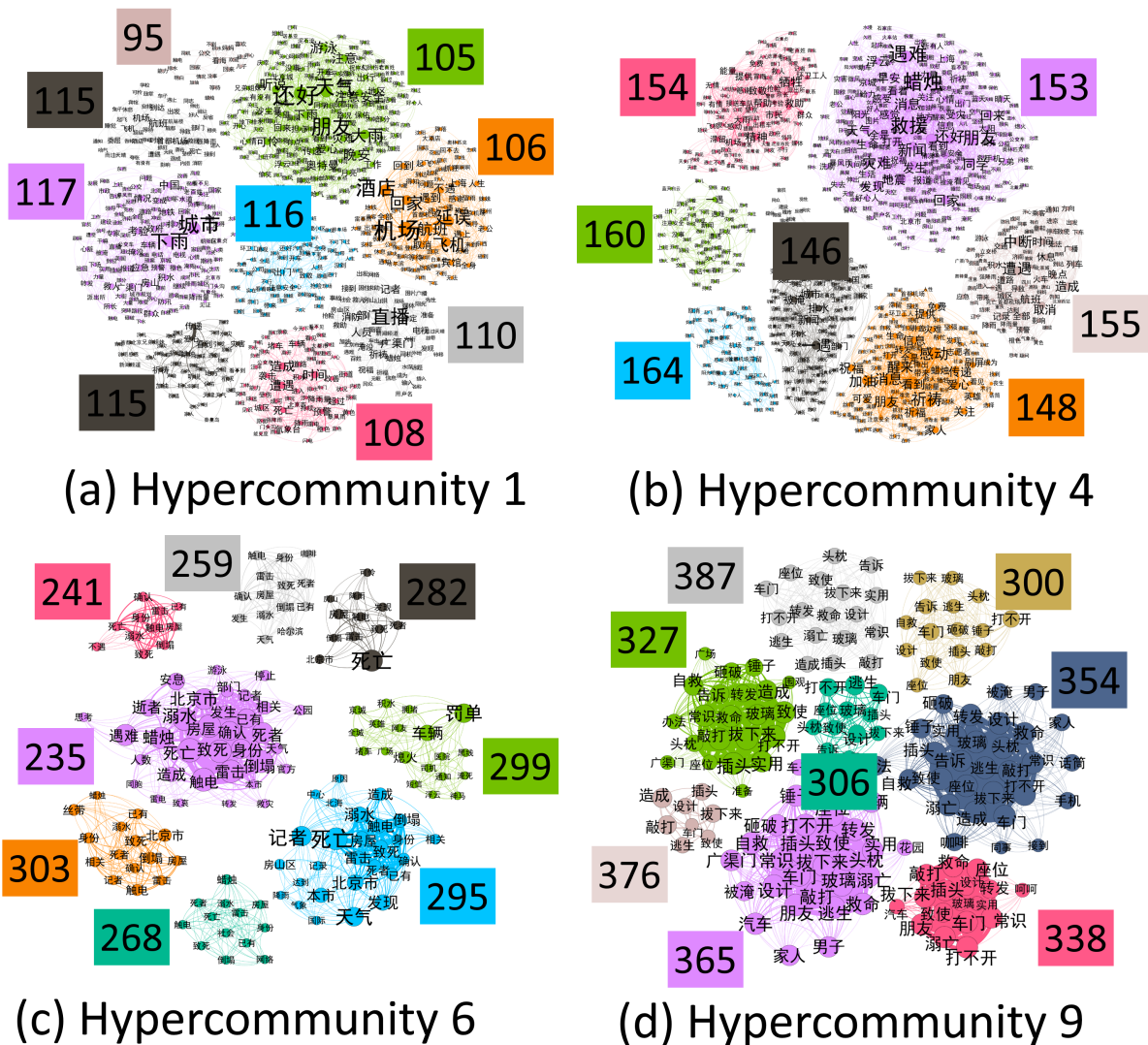


FIGURE 10: Main co-word network communities (top10 in community size, and the community size is larger than 3) corresponding to the nodes of four sample hypercommunities. Each co-word network community is labeled by a unique number corresponding to the unique number of the nodes of hypercommunities in Fig. 9. Only the main nodes of each community (when nodes in the head of node degree distribution according to head/tail break) are shown in this figure.

TABLE 4: Topics and topology of each hypercommunities.

Hypercommunity ID	Topics	topology
1	various	network
2	flight delays and cancellation	tree
3	joke on waterlogging	line
4	various (primarily regarding friends and news)	network
5	various (primarily regarding discussions and accusations against the government)	network
6	death	tree
7	various (primarily regarding prayer for casualties and discussions regarding urban construction)	network
8	joke on waterlogging	line
9	self-seeking method	line
10	various (primarily regarding reflection on disaster rationality)	network

primarily included the following: flight delays and cancellations; reminding friends of safety; discussion regarding urban facilities, casualties, and prayers. Hypercommunity 4 is also formed by various topics. Its main topics are friends and news.

VII. DISCUSSION

A. DISCOVERING TOPICS FOR SHORT TEXTS

In this study, we demonstrate that the community detection of the co-word network can be regarded as a general topic model, which can be used to discover topics for short texts. We identified the topic words in Weibo messages (short texts) based on the TF-IDF and head/tail break method and constructed the co-word network according to the co-occurrence of these topic words. Because the community typically refers to a closely connected group in a network, the community of the co-word network can be regarded as a specific topic. The topic can be identified by examining the topic words in the community.

B. SETTING THE THRESHOLD OF INCLUSION

We used a simplified version of the GED method [47] to identify topic evolution events based on the co-word network in social media. The inclusion-based GED method is a simple method for identifying community evolution events. However, in practice, methods to set the appropriate threshold of the inclusion measure are lacking. In this study, we provide a new method to set the threshold of inclusion by observing the distribution of the inclusion and using the head/tail break method [57]. This method can be used as a reference for the study of community evolution of the co-word network.

C. QUANTIFYING THE EVOLUTION OF TOPICS

We herein proposed an evolutionary community network, a flexible method, to support and quantify the evolution of topics for short text. The community evolutionary network is a directed network that can be further analyzed using network analysis methods, such as degree analysis and community analysis. We used the community analysis method as an example to quantify the evolution of topics during the rainstorm and focused on the time and semantics of the community. The hypercommunity of the community evolutionary network reveals the specific local evolution process like a magnifying glass, and the lifetime of the hypercommunity reveals the process and stage of the disaster.

VIII. CONCLUSION

Situational awareness refers to the acquisition of context-related information in an event, which is important for disaster management. In this study, we proposed a method for quantifying the evolution of topics based on social media texts to enhance situational awareness in disasters. In a case study of “July 2012 Beijing flood,” we discovered that the communities of the co-word network could be regarded as topics in social media messages during an event. Furthermore, network analysis methods based on the community

evolutional network could be used to quantify the evolution of topics. Additionally, we discovered a corresponding relationship between the topic evolution and the rainstorm stage. Emergency managers can use the method to analyze the process and stages of disasters based on social media.

Our approach presents some limitations and therefore further research is required. Using the community detection of the co-word network to discover topics often neglect the semantics of the topic words. Two topic words may be semantically similar but not appear in the same text, which may not be correctly divided into the same topic. In future studies, we will use the semantic similarity between topic words to adjust the edge weight of the co-word network to obtain more reasonable topic communities. We applied the Louvain method for community detection on the constructed co-word network. Other community detection algorithms can also be used in our framework to detect topics and topic evolution. For example, the overlapping community detection algorithm can be used to identify the overlapping of topics. The corresponding work can be discussed in future work. In addition, we will investigate a more flexible method to set the length of time windows according to the actual data in different disasters. Finally, complex hypercommunities can be used to conduct second community detection to obtain a more localized topic evolution.

REFERENCES

- [1] Q. Huang and Y. Xiao, “Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery,” *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 3, pp. 1549–1568, Aug. 2015.
- [2] Y.-d. Wang, X.-k. Fu, W. Jiang, T. Wang, M.-H. Tsou, and X.-y. Ye, “Inferring urban air quality based on social media,” *Computers, Environment and Urban Systems*, vol. 66, no. Supplement C, pp. 110–116, Nov. 2017.
- [3] Crooks Andrew, Croitoru Arie, Stefanidis Anthony, and Radzikowski Jacek, “#Earthquake: Twitter as a Distributed Sensor System,” *Transactions in GIS*, vol. 17, no. 1, pp. 124–147, Oct. 2012.
- [4] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, “Using Social Media to Enhance Emergency Situation Awareness,” *IEEE Intell. Syst.*, vol. 27, no. 6, pp. 52–59, Nov. 2012.
- [5] J. Li, Z. He, J. Plaza, S. Li, J. Chen, H. Wu, Y. Wang, and Y. Liu, “Social Media: New Perspectives to Improve Remote Sensing for Emergency Response,” *Proc. IEEE*, vol. 105, no. 10, pp. 1900–1912, Oct. 2017.
- [6] A. Kongthong, C. Haruechaiyasak, J. Pailai, and S. Kongyoung, “The role of Twitter during a natural disaster: Case study of 2011 Thai Flood,” in *2012 Proceedings of PICMET '12: Technology Management for Emerging Technologies*, Vancouver, BC, Canada, Jul. 2012, pp. 2227–2232.
- [7] Y. Wang, T. Wang, X. Ye, J. Zhu, and J. Lee, “Using Social Media for Emergency Response and Urban Sustainability: A Case Study of the 2012 Beijing Rainstorm,” *Sustainability*, vol. 8, no. 1, p. 25, Dec. 2015.
- [8] Q. Deng, G. Cai, H. Zhang, Y. Liu, L. Huang, and F. Sun, “Enhancing Situation Awareness of Public Safety Events by Visualizing Topic Evolution Using Social Media,” in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, ser. Dg.o '18. Delft, The Netherlands: ACM, 2018, pp. 7:1–7:10.
- [9] L. Ferrari, A. Rosi, M. Mamei, and F. Zambonelli, “Extracting urban patterns from location-based social networks,” in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks - LBSN '11*. Chicago, Illinois: ACM Press, 2011, p. 1.

- [10] M.-H. Tsou, J.-A. Yang, D. Lusher, S. Han, B. Spitzberg, J. M. Gawron, D. Gupta, and L. An, "Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): A case study in 2012 US Presidential Election," *Cartogr. Geogr. Inf. Sci.*, vol. 40, no. 4, pp. 337–348, Sep. 2013.
- [11] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welp, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," in *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*, 2010.
- [12] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *Int. J. Geogr. Inf. Sci.*, vol. 0, no. 0, pp. 1–24, Aug. 2017.
- [13] S. Steensen, E. Frey, H. Hornmoen, R. Ottosen, and M. T. Konow-Lund, "Social Media and Situation Awareness during Terrorist Attacks: Recommendations for Crisis Communication," in *Social Media Use in Crisis and Risk Communication*. England, United Kingdom: Emerald Publishing Limited, Sep. 2018, pp. 277–295.
- [14] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," in *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*. Atlanta, Georgia, USA: ACM Press, 2010, p. 1079.
- [15] B. Lindsay, "Social Media and Disasters: Current Uses, Future Options, and Policy Considerations," *J Curr Issues Media Telecommun*, vol. 2, pp. 1–14, Jan. 2012.
- [16] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Hum Factors*, vol. 37, no. 1, pp. 32–64, Mar. 1995.
- [17] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford, "SensePlace2: GeoTwitter analytics support for situational awareness," in 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, Rhode Island, USA, Oct. 2011, pp. 181–190.
- [18] Q. Huang, G. Cervone, D. Jing, and C. Chang, "DisasterMapper: A CyberGIS Framework for Disaster Management Using Social Media Data," in *Proceedings of the 4th International ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data*, ser. BigSpatial'15. Bellevue, WA, USA: ACM, 2015, pp. 1–6.
- [19] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. V. Hentenryck, J. Fowler, and M. Cebrian, "Rapid assessment of disaster damage using social media activity," *Sci. Adv.*, vol. 2, no. 3, p. e1500779, Mar. 2016.
- [20] Y. Wang, S. Ruan, T. Wang, and M. Qiao, "Rapid estimation of an earthquake impact area using a spatial logistic growth model based on social media data," *Int. J. Digit. Earth*, vol. 0, no. 0, pp. 1–20, Jul. 2018.
- [21] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. Raleigh, North Carolina, USA: ACM, 2010, pp. 851–860.
- [22] V. Slavkovikj, S. Verstockt, S. Van Hoecke, and R. Van de Walle, "Review of wildfire detection using social media," *Fire Safety Journal*, vol. 68, pp. 109–118, Aug. 2014.
- [23] H. Zhang, G. Kim, and E. P. Xing, "Dynamic Topic Modeling for Monitoring Market Competition from Online Text and Image Data," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. Sydney, NSW, Australia: ACM Press, 2015, pp. 1425–1434.
- [24] X. Wang and A. McCallum, "Topics over time: A non-Markov continuous-time model of topical trends," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Philadelphia, PA, USA: ACM, 2006, pp. 424–433.
- [25] L. AlSumait, D. Barbará, and C. Domeniconi, "On-line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking," in 2008 Eighth IEEE International Conference on Data Mining. Pisa, Italy: IEEE, Dec. 2008, pp. 3–12.
- [26] M. Hoffman, F. R. Bach, and D. M. Blei, "Online learning for latent dirichlet allocation," in *Advances in Neural Information Processing Systems*, Vancouver, British Columbia, Canada, 2010, pp. 856–864.
- [27] B. Yu, L. Wang, and W. Zhang, "Topic evolution analysis based on dual-OLDA model under Chinese semantic environment," in 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, Mar. 2017, pp. 658–664.
- [28] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda, "Topic Tracking Model for Analyzing Consumer Purchase Behavior," in *IJCAI*, Pasadena, California, USA, 2009.
- [29] R. M. Nallapati, S. Dittmore, J. D. Lafferty, and K. Ung, "Multiscale topic tomography," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '07*. San Jose, California, USA: ACM Press, 2007, p. 520.
- [30] J. Zhu, X. Li, M. Peng, J. Huang, T. Qian, J. Huang, J. Liu, R. Hong, and P. Liu, "Coherent topic hierarchy: A strategy for topic evolutionary analysis on microblog feeds," in *International Conference on Web-Age Information Management*. Qingdao, China: Springer, 2015, pp. 70–82.
- [31] M. García Lozano, J. Schreiber, and J. Brynielsson, "Tracking geographical locations using a geo-aware topic model for analyzing social media data," *Decision Support Systems*, vol. 99, pp. 18–29, Jul. 2017.
- [32] D. Rind, "Complexity and Climate," *Science*, vol. 284, no. 5411, pp. 105–107, Apr. 1999.
- [33] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, p. 026113, Feb. 2004.
- [34] H.-J. Li, Z. Bu, Z. Wang, and J. Cao, "Dynamical clustering in electronic commerce systems via optimization and leadership expansion," *IEEE Trans. Ind. Inf.*, pp. 1–1, 2019.
- [35] J. Cao, Z. Bu, Y. Wang, H. Yang, J. Jiang, and H.-J. Li, "Detecting Prosumer-Community Groups in Smart Grids From the Multiagent Perspective," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 49, no. 8, pp. 1652–1664, Aug. 2019.
- [36] H.-J. Li, Q. Wang, S. Liu, and J. Hu, "Exploring the trust management mechanism in self-organizing complex network based on game theory," *Physica A: Statistical Mechanics and its Applications*, p. 123514, Nov. 2019.
- [37] Z. Bu, H.-J. Li, C. Zhang, J. Cao, A. Li, and Y. Shi, "Graph K-means based on Leader Identification, Dynamic Game and Opinion Dynamics," *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2019.
- [38] H.-J. Li and J. J. Daniels, "Social significance of community structure: Statistical view," *Phys. Rev. E*, vol. 91, no. 1, p. 012801, Jan. 2015.
- [39] Y.-Q. Zhang and X. Li, "Temporal dynamics and impact of event interactions in cyber-social populations," *Chaos*, vol. 23, no. 1, p. 013131, Feb. 2013.
- [40] K. Hu, K. Qi, Q. Guan, C. Wu, J. Yu, Y. Qing, J. Zheng, H. Wu, and X. Li, "A Scientometric Visualization Analysis for Night-Time Light Remote Sensing Research from 1991 to 2016," *Remote Sens.*, vol. 9, no. 8, p. 802, Aug. 2017.
- [41] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig, "Community detection in networks: A multidisciplinary review," *Journal of Network and Computer Applications*, vol. 108, pp. 87–111, Apr. 2018.
- [42] Y. Zhang, L. Wang, Y.-Q. Zhang, and X. Li, "Towards a temporal network analysis of interactive WiFi users," *EPL*, vol. 98, no. 6, p. 68002, Jun. 2012.
- [43] S. Saganowski, P. Bródka, and P. Kazienko, "Community Evolution," *ArXiv Prepr. ArXiv160500069*, 2016.
- [44] H. Xu, Y. Hu, Z. Wang, J. Ma, and W. Xiao, "Core-Based Dynamic Community Detection in Mobile Social Networks," *Entropy*, vol. 15, no. 12, pp. 5419–5438, Dec. 2013.
- [45] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Facetnet: A framework for analyzing communities and their evolutions in dynamic networks," in *Proceeding of the 17th International Conference on World Wide Web - WWW '08*. Beijing, China: ACM Press, 2008, p. 685.
- [46] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu, "GraphScope: Parameter-free mining of large time-evolving graphs," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '07*. San Jose, California, USA: ACM Press, 2007, p. 687.
- [47] P. Bródka, S. Saganowski, and P. Kazienko, "GED: The method for group evolution discovery in social networks," *Soc. Netw. Anal. Min.*, vol. 3, no. 1, pp. 1–14, Mar. 2013.
- [48] W. Ding and C. Chen, "Dynamic topic detection and tracking: A comparison of HDP, C-word, and cocitation methods," *Journal of the Association for Information Science and Technology*, vol. 65, no. 10, pp. 2084–2097, Oct. 2014.
- [49] Z. Chen and X. Sun, "Evolution Analysis of Topics on Social Media Based on the Co-word Network," in 2016 International Seminar on Education Innovation and Economic Management (SEIEM 2016). Chongqing, China: Atlantis Press, 2016.
- [50] C. Fan, A. Mostafavi, A. Gupta, and C. Zhang, "A System Analytics Framework for Detecting Infrastructure-Related Topics in Disas-

ters Using Social Sensing,” in EG-ICE 2018: Advanced Computing Strategies for Engineering, ser. Lecture Notes in Computer Science, I. F. C. Smith and B. Domer, Eds., Lausanne, Switzerland: Springer International Publishing, 2018, pp. 74–91.

- [51] S. Gründer-Fahrer, A. Schlaf, G. Wiedemann, and G. Heyer, “Topics and topical phases in German social media communication during a disaster,” *Nat. Lang. Eng.*, vol. 24, no. 2, pp. 221–264, Mar. 2018.
- [52] A. Rajaraman and J. D. Ullman, “Data Mining,” in *Mining of Massive Datasets*. Cambridge, England, United Kingdom: Cambridge University Press, 2011, pp. 1–17.
- [53] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *J. Stat. Mech.*, vol. 2008, no. 10, p. P10008, 2008.
- [54] S. Fortunato and M. Barthélemy, “Resolution limit in community detection,” *PNAS*, vol. 104, no. 1, pp. 36–41, Jan. 2007.
- [55] “July 2012 Beijing flood,” Wikipedia, Sep. 2019, page Version ID: 918807586.
- [56] J. SUN, N. HE, G. WANG, M. CHEN, X. LIAO, and H. WANG, “Preliminary analysis on synoptic configuration evolvement and mechanism of a torrential rain occurring in Beijing on 21 July 2012,” *Torrential Rain Disasters*, vol. 31, no. 3, pp. 218–225, 2012.
- [57] B. Jiang, “Head/tail Breaks: A New Classification Scheme for Data with a Heavy-tailed Distribution,” *Prof. Geogr.*, vol. 65, no. 3, pp. 482–494, Aug. 2013.
- [58] R. H. Baayen, *Word Frequency Distributions*, ser. Text, Speech and Language Technology. Heidelberg, Germany: Springer Netherlands, 2001.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [60] S. van der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy Array: A Structure for Efficient Numerical Computation,” *Comput. Sci. Eng.*, vol. 13, no. 2, pp. 22–30, Mar. 2011.
- [61] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using NetworkX,” in *Proceedings of the 7th Python in Science Conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11–15.
- [62] C. L. Staudt, A. Sazonovs, and H. Meyerhenke, “NetworKit: A Tool Suite for Large-scale Complex Network Analysis,” *ArXiv14033005 Phys.*, Nov. 2015.
- [63] “Sina Weibo,” Wikipedia, Jan. 2020, page Version ID: 934191705.
- [64] “2012 Yangzhou earthquake,” Wikipedia, Dec. 2018, page Version ID: 872698309.



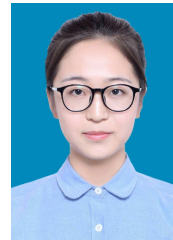
XIAOKANG FU was born in Xiangcheng, Henan province, China in 1991. He received the B.S. degree in Engineering of Surveying and Mapping from Inner Mongolia University of Science&Technology, Baotou, in 2013 and the M.S. degree in Engineering of Surveying and Mapping from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, in 2015. He is currently a Ph.D. Candidate, in Car-

tography and Geographic Information Engineering from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China.

His research interests include remote sensing, geostatistics and geoinformatics (GIS), urban computing, social media data mining, machine learning, and emergency response.



YANDONG WANG is currently a Professor is in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interest mainly concentrates in space information intelligent service, GIS theory and engineering application, and spatio-temporal information knowledge mining.



MENGMENG LI was born in Changzhi, Shanxi in 1995. She received the B.S. degree in geographic information science from Dalian Maritime University, Dalian, Liaoning, in 2017. She is currently working toward the M.S. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.



MINGXUAN DOU received the M.S. degree from Wuhan University, Wuhan, China, in 2018. She is currently working toward the PH.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University.



MENGLING QIAO is currently working toward the PH.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. Her research interests include spatiotemporal data mining, and urban analysis.



agement.

KAI HU received the Bachelor degree and Master degree from Wuhan University in 2011 and 2013, respectively. He received the doctor degree of Cartography and Geographical Information Engineering from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University in 2017. He is currently working in Jiangnan University as associate professor. His research interest include text mining, scientometric, and information man-

...