

# Information Diffusion on Social Media During Natural Disasters

Rongsheng Dong, Libing Li, Qingpeng Zhang<sup>✉</sup>, *Member, IEEE*, and Guoyong Cai

**Abstract**—Social media analytics has drawn new quantitative insights of human activity patterns. Many applications of social media analytics, from pandemic prediction to earthquake response, require an in-depth understanding of how these patterns change when human encounter unfamiliar conditions. In this paper, we select two earthquakes in China as the social context in Sina-Weibo (or Weibo for short), the largest Chinese microblog site. After proposing a formalized Weibo information flow model to represent the information spread on Weibo, we study the information spread from three main perspectives: individual characteristics, the types of social relationships between interactive participants, and the topology of real interaction networks. The quantitative analyses draw the following conclusions. First, the shadow of Dunbar's number is evident in the "declared friends/followers" distributions, and the number of each participant's friends/followers who also participated in the earthquake information dissemination show the typical power-law distribution, indicating a rich-gets-richer phenomenon. Second, an individual's number of followers is the most critical factor in user influence. Strangers are very important forces for disseminating real-time news after an earthquake. Third, two types of real interaction networks share the scale-free and small-world property, but with a looser organizational structure. In addition, correlations between different influence groups indicate that when compared with other online social media, the discussion on Weibo is mainly dominated and influenced by verified users.

**Index Terms**—Emergency response, online collective behavior, Sina-Weibo, social network analysis.

## I. INTRODUCTION

THE 2009 DARPA Red Balloon Challenge offered the Internet and social networking a chance to demonstrate their vast potential to solve a distributed, time-critical, highly distributed public problem [1], [2]. And the modern social network study has inferred much new quantitative knowledge

about human activity patterns, such as influencer's identification [2]–[5], the network topology measurement [6], [7], trust analysis [8]–[10], social hot spot-tracing [11]–[13], and the dynamics of information spread [14]–[16]. However, many applications, from pandemic prediction to earthquake response, require an understanding of how these patterns change when human encounter unfamiliar conditions [17], [18]. Especially for China, who suffered from frequent natural disasters, the understanding of how the behaviors of hundreds of millions of Web users change is very important. The empirical study of the human flesh search (HFS) [19], [20], for one, provided quantitative insights into these collective responses of Web users in China. Inspired by previous research on Web users collective responses, we choose two empirical cases in China-Yi'liang (2012) and Ya'an earthquakes (2013). What makes it additionally useful is that many densely populated areas in mainland China, such as Sichuan, Fujian, and so on, are frequent earthquake areas and often suffered severe damage from earthquakes.

At present for earthquake topics, there are two main types of studies: detecting seismic waves and enhancing rescue efforts. The former focuses on how to improve the accuracy of magnitude of earthquake forecasting or issue warnings as early as possible [21]–[24], such as the Did You Feel It system. And the later studies [25]–[27] try to explore ideas to cope with earthquake relief, postearthquake reconstruction, and to improve the mental health status of rescuers. This paper focuses on the latter effort from a social network perspective, with a particular focus on information diffusion and social networking behaviors.

In social network study, the above-mentioned focus [25]–[27] can be seen as a nonlinear superposition of a multitude of social interaction networks, where nodes represent individuals and edges capture a variety of different social relations. However, after further study, a group of researchers represented by Huberman *et al.* [28] found that social interactions within Twitter cannot be inferred directly from a declared relationship set of friends and followers: many users interact with few other people in their declared relationship network [28]–[31]. The key problem is that the structure of the underlying interaction network is not visible and must be inferred from the flow of information between individuals, which poses a serious challenge to our efforts to understand how the structure of the network affects social dynamics and the spread of information [32].

Take the tweet reposting connection shown in Fig. 1(a) and (b) as examples. There are four features in the

Manuscript received May 15, 2017; revised December 2, 2017; accepted December 5, 2017. Date of publication January 11, 2018; date of current version February 23, 2018. This work was supported in part by the National Natural Science Foundation of China under 61762024, Grant 61763007, Grant 71402157, Grant 71672163, Grant 61363070, Grant 61063039, and Grant 61262074, in part by the Natural Science Foundation of Guangxi Province under Grant 2017GXNSFDA198050, and in part by the Natural Science Foundation of Guangdong Province under Grant 2014A030313753. (Corresponding author: Qingpeng Zhang.)

R. Dong and G. Cai are with the Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: ccrsdong@guet.edu.cn; ccgycai@guet.edu.cn).

L. Li is with the Mintian Insurance Surveyors & Loss Adjusters Group Company Ltd., Shenzhen 518000, China (e-mail: lilibing1@163.com).

Q. Zhang is with the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, and also with the Shenzhen Research Institute of City University of Hong Kong, Shenzhen 518000, China (e-mail: qingpeng.zhang@cityu.edu.hk).

Digital Object Identifier 10.1109/TCSS.2017.2786545

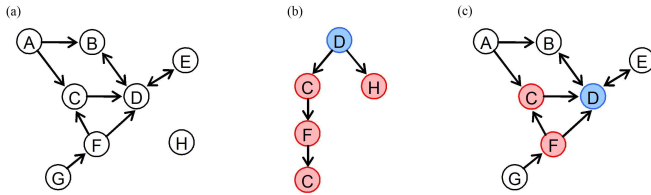


Fig. 1. Typical multirelational participant network on Twitter/Weibo. (a) Declared social network, where nodes represent users and directed edges represent relationships of followers or friends. (b) Information cascade of one tweet, where nodes represent participating users and directed edges indicate tweets citing relationships. (c) Typical multirelational participant network, where the blue node is the original poster, pink nodes are the reposters, and edges still indicate user relationships.

underlying interaction network: 1) user-D has four followers, B, C, F, and E, but just C and F repost the message created by D; 2) user-F reposts the tweet created by user-D through the intermediary-C; 3) user-C participates more than once; and 4) the tweet created by user-D is also reposted by user-H who does not have the relationship of friend or follower with user-D.

There are two ways of describing these features in previous studies. The conventional method distinguishes whether users interact with each other by adding extra attributes for nodes or edges of declared relationship networks [17], [18], [33]. Fig. 1(c) is the typical application of this method, which attempts to describe Fig. 1(a) and (b) together. However, it cannot depict the features-III and IV. The other solution is to describe all kinds of user relationships or interactions using a multilayer/multirelational network [34], [35]. Likewise, the method cannot express the features-II and III explicitly, and the types of models lack the parallel development of specific analysis methods to exploit the information hidden between the layers.

In response to the above-mentioned problem, we propose a formalized definition of the Weibo information flow (WIF) and have applied it to the empirical analysis of a Sina-Weibo data set on the topic of earthquakes. The goal of our work is to provide a way to extract the underlying hidden social network that more clearly represents the information cascade and actual interactions among users after an earthquake. We will then address the problem in understanding of hidden social networks behind the online information flow.

- 1) What about the languages used by participants? How is the demographics of participant users? Are there regional features? What are the differences before and after the earthquake?
- 2) What are the common features of influential users? What are the key factors in user influence?
- 3) When a user reposts, will he/she evenly repost tweets that attract him/her, or just those tweets posted by his/her friends? Also for users whose message are reposted, does the reposting all take place solely among their followers? Are there any differences in information diffusion patterns before and after the earthquake?
- 4) Compared with early Web2.0 online communications, does the Sina-Weibo platform have unique features that include social and information dissemination functions?

- 5) Are there any correlations between the type of interacted user relationships and the topological feature of user networks?

The organization of this paper is as follows. Section III presents the main body of this paper. We first introduce the data set and give a formal definition of the WIF model in Section II. Section III consists of three subsections. Section III-A includes the empirical results of eight individual attributes: the configured language of a personal page, gender, location, the number of followers, the number of friends, the number of posted tweets, the number being reposted, and the correlation among three user rankings. Section III-B analyzes the proportion of each type of user relationship between interacted users and discusses the differences in reposting patterns of each interacted user group by the user relationships. Section III-C uses social network analysis to unveil the topological properties of two types of underlying interaction networks, which are extracted by the WIF model, and analyze the correlation between influential user groups of two networks. Section IV closes this paper with remarks for future work.

## II. DATA SET AND METHODOLOGIES

### A. Data Set

Yi'liang earthquake erupted on September 7, 2012. Ya'an earthquake erupted on April 20, 2013. Because happening consecutively in China, these two natural disasters triggered sparked discussions of Sina-Weibo users. We collected these tweet repost data related two disasters by the Sina-Weibo public application programming interface (API). The API's registered users can obtain relevant permits as long as that their identities are verified by the Sina-Weibo user authentication. During the course of empirical analyses, we used the MySQL database management system for data extraction and cleansing, and used the Cytoscape toolkit to analyze the network topology [36]. All of the figures are plotted with MATLAB.

As presented in Table I, the data set consists of two parts, demonstrating the contrast between before and after earthquake. Yi'liang county in China was hit by a 5.7-magnitude earthquake on September 7, 2012; the section titled *After Yi'liang Earthquake* is the seismic information generated by Sina-Weibo for seven months after the earthquake. In addition, the section *Before Ya'an Earthquake* is the daily information exchanged on Sina-Weibo regarding Ya'an within a week before April 20, 2013, a 7.0-magnitude earthquake occurred. The object of comparison should be the "before Yi'liang earthquake," but the amount of relevant data is too small for empirical analysis.

Each original tweet record consists of five parts: the original tweet, the original user, repost tweets and their corresponding users, the repost tree, and the relationship between users. Fig. 2 shows the variation trend over time of the total number of these tweets and retweets and five high peaks in the after Yi'liang earthquake data set. These peaks may be related to some sensitive events taking place during those periods. The following social events might offer some available clues.

TABLE I  
SINA-WEIBO DATA SET

Social context of the Dataset	$NUM_{OT}$	$NUM_{RT}$	$NUM_{WU}$	$NUM_{TRT}$	Time span
After Yi'liang Earthquake	9396	407584	330466	3096	Sep 7, 2012 - Apr 30, 2013
Before Ya'an Earthquake	1918	49416	46682	811	Apr 13, 2013 - Apr 19, 2013

$NUM_{OT}$ : The count of original tweets;  $NUM_{RT}$ : The count of retweets;  $NUM_{WU}$ : The count of Sina-Weibo users;  $NUM_{TRT}$ : The count of tweet repost trees (A tweet repost tree consists of one unique original tweet, it's retweets and repost relationships among them.).

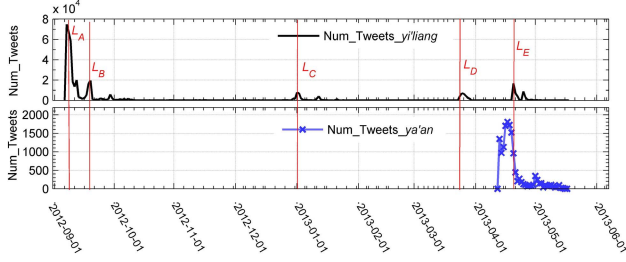


Fig. 2. Change in the total number of related original tweets and their retweets over time. Num\_Tweets\_Yi'liang: number of tweets created after the Yi'liang earthquake, Num\_Tweets\_Ya'an: number of tweets created before the Ya'an earthquake. Five vertical red lines titled  $L_A$ ,  $L_B$ ,  $L_C$ ,  $L_D$ , and  $L_E$  represent the five dates when important events took place.

- 1)  $L_A$ -September 9, 2012: A 5.7-magnitude earthquake occurred in Yiliang county of China.
- 2)  $L_B$ -September 19, 2012: The Yiliang county published the latest donations list for Yiliang earthquake.
- 3)  $L_C$ -December 30, 2012: Called the kindest substitute teacher, Zhu Yinquan accidentally fell from a building, which resulted in severe brain injury, and he had saved seven students by hand in Yiliang earthquake.
- 4)  $L_D$ -March 24, 2013: The news of relief fund was not distributed in time flashed on the Internet.
- 5)  $L_E$ -April 20, 2013: A 7.0-magnitude earthquake occurred in Ya'an of Sichuan province, China.

### B. Weibo Information Flow (WIF) Model

To describe and retrieve a Weibo-spread information precisely, a formal definition of WIF is proposed as the following quadruple:

$$WIF = \langle WUS, TS, RRS, TRTS \rangle$$

where  $WUS$  is the Weibo user set consisting of participant Weibo users,  $TS$  is the tweet set consisting of original tweets and retweets,  $RRS$  is the repost relationship set consisting of the citing relationships between tweets, and  $TRTS$  is the tweet reposts tree set consisting of repost trees of each original tweet.

#### 1) Structure of $WU(\in WUS)$ :

$$WU = \langle gender, pNum, lang, followerSet, friendSet, sCount \rangle$$

where  $gender = 0/1$  is the "0" to male and "1" to female.  $pNum \in ProvincesNumList$ ,  $pNum$  denotes the city where the current user lives (see the Sina-Weibo city code list).  $lang \in \{zh-cn, zh-tw, zh-hk, en\}$  represents the language that the current user has configured on his Weibo page, where "zh-cn"

denotes the simplified Chinese characters used in Mainland China, "zh-tw" and "zh-hk" denotes the complex Chinese characters used in Taiwan and Hong Kong, respectively, and "en" denotes English.  $followerSet$  is the set consisting of Weibo users who follow the current user. For the  $followerSet/friendSet$  of each participant, the actual data set mentioned in Table I just records a number, i.e., the number of his/her followers/friends. Of course, for the part of their followers/friends who also participated in the earthquake information dissemination by posting/reposting Sina-Weibo messages, there are detailed records in the data set.  $friendSet$  is the set consists of Weibo users whom the current user follows.  $sCount$  is the number of tweets posted by the current user.

#### 2) Structure of $T(\in TS)$ :

$$T = \langle flag, user, repostCount, time, text, ot \rangle$$

where  $flag = 0/1$ , "0" indicates that the current tweet is an original tweet and "1" indicates responding to the retweet,  $user \in WUS$  is the user who posted the current tweet,  $repostCount \in N$  is the number of times of the current tweet has been reposted,  $time$  is the time when the current tweet is posted,  $text$  is the content body of the current tweet, and  $ot \in TS$ , when  $flag = 1$ , it denotes the original tweet of the current tweet; otherwise, it is null.

#### 3) Structure of $RR(\in RRS)$ :

$$RR = \langle st, rt, depth, type \rangle$$

where  $st \in \{T \mid (T \in TS) \wedge (TrepostCount > 0)\}$  denotes the tweet that was reposted. Regular occurrences of the multilayer reposting phenomenon exist: for example, user-A posts the  $ot$  first, and then user-B posts the  $rt_B$  by reposting  $ot$ , and  $rt_B$  are reposted by others, too. Therefore, the  $st$  is either an original tweet or a retweet.  $rt \in \{T \mid (T \in TS) \wedge (T.flag = 1)\}$  is the retweet generated by reposting  $st$ .  $depth \in N^+$ , indicates the distance between  $rt$  and  $rt.ot$  in the repost tree where  $N^+ = \{1, 2, 3, \dots\}$ , and if  $RR.st.flag = 0$ , then  $RR.depth = 1$ .  $type \in \{I, II, III, IV, V\}$ , denotes five repost-types by relationship types between the  $st.user$  and the  $rt.user$ . Fig. 3(a) shows three types of user relationships among Weibo users where the directed edge indicates that the starting user follows the target user. Here, we refer to the pair of users linked by bidirectional edges as *bi-friend* if they both follow each other, and refer those users as *strangers* if there are not any directed edges. In this paper, it is assumed that if user-B reposts one tweet posted by user-A, then the tweet information flows from user-A to user-B. After the combination of the information direction between two users and their relationship type, there will be five types



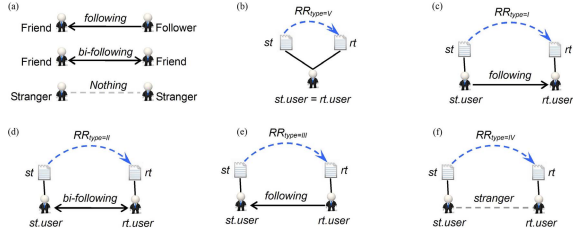


Fig. 3. Three types of user relationships and five types of information cascade patterns in WIF. (a) Three types of Weibo-user relationships. (b)  $RR$  where  $type = V$ . (c)  $RR$  where  $type = I$ . (d)  $RR$  where  $type = II$ . (e)  $RR$  where  $type = III$ . (f)  $RR$  where  $type = IV$ .

of  $RR$ s [see Fig. 3(b)–(f)]. The following equations are the symbolic descriptions of five types of  $RR$ s:

$RRtype$

$$= V \iff RR.st.user = RR.rt.user$$

$RR.type$

$$= I \iff (RR.st.user \in RR.rt.user.followerset) \wedge (RR.rt.user \notin RR.st.user.followerset)$$

$RR.type$

$$= II \iff (RR.st.user \in RR.rt.user.followerset) \wedge (RR.rt.user \in RR.st.user.followerset)$$

$RR.type$

$$= III \iff (RR.st.user \notin RR.rt.user.followerset) \wedge (RR.rt.user \in RR.st.user.followerset)$$

$RR.type$

$$= IV \iff (RR.st.user \notin RR.rt.user.followerset) \wedge (RR.rt.user \notin RR.st.user.followerset).$$

4) Structure of  $TRT(\in TRTS)$ :

$$TRT = \langle OT, RTS, RRS^*, size \rangle$$

where  $OT \in \{T \mid (T \in TS) \wedge (T.flag = 0)\}$  indicates the original tweet represented by the root node of the current tweet repost tree.  $RTS^* \subseteq \{T \mid (T \in TS) \wedge (T.flag = 0) \wedge (T.ot = OT)\}$  indicates the node set of the current  $TRT$ , representing  $RT$ s generated by reposting  $OT$ .  $RRS^* \subseteq \{RR \mid (RR \in RRS) \wedge (RR.RT \in RRS^*)\}$  indicates the edge set of the current  $TRT$ , representing  $RR$ s generated by reposting  $OT$ .  $size = (OT.repostCount + 1)$  indicates the number of nodes of the current  $TRT$ .

### III. RESULTS AND DISCUSSION

#### A. Analysis of Users' Characteristics

This subsection focuses on the following questions. What about the distribution of users' language? How is demographics of participant users? Are there some regional features? What are the differences before and after the earthquake? What are the typical features of influential users?

1) *Basic Analysis*: The demographics, language, and geographic distribution of online users are important measurement indicators of application fields in disaster relief and disease surveillance and control [21], [37], [38]. On Sina-Weibo,

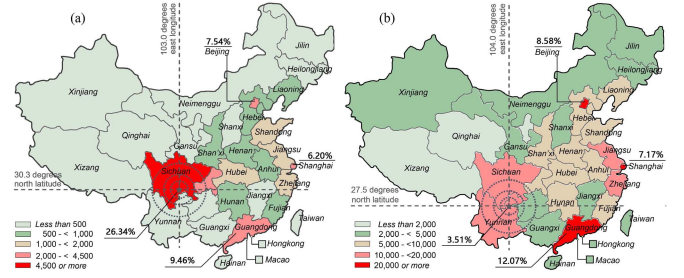


Fig. 4. Geographic distributions of Sina-Weibo participants before and after earthquakes in China by province. Geographic distribution of participants (a) before Ya'an earthquake and (b) after Yi'liang earthquake.

the growth spurt of seismic topic tweets emerged as soon as the earthquake occurred. Where did these active users come from? What are their social backgrounds? Our measurement of  $WU.gender$ ,  $WU.lang$ , and  $WU.pNum$  shows: 1) there is a roughly equal proportion of males and females among participant users; 2) users living in the Chinese Mainland, Hong Kong, Macao, and Taiwan are the force that made the tweet peak instantaneously, and the reason of the phenomena mainly come from the locality of Sina-Weibo; and 3) there are a few foreign users participating to post or repost related tweets, but nearly all of them live in China.

Fig. 4 shows the geographic distribution of participant users before and after the earthquake. During the aftermath in Yi'liang which was hit by a 5.7-magnitude earthquake on September 7, 2012, Fig. 4(b) shows that Beijing, Shanghai, and Guangdong topped the rank in the number of participant users and the proportion of active users located in these three areas is 27.83%, these three cities are the political center, the financial center, and the largest province in the economy, respectively. Only 3.5% of active users are located in the local province (Yunnan). In cases prior to the earthquake, there is too little data to analyze unfortunately, probably because Yi'liang is a very remote country in Yunnan province. Therefore, we replaced it with Ya'an which is a popular tourist destination (in the same way later). Then, we found in Fig. 4(a) that local province (Sichuan) ranked first in the number of active users posting Ya'an common topics, with a share of 26.33%, at the same time the sum of shares in Beijing, Shanghai, and Guangdong still reaches 29.25%.

For participants after the Yi'liang earthquake, Fig. 5 shows the distributions of their four individual attributes the number of friends ( $WU.frCount$ ), the number of followers ( $WU.frCount$ ), and the number of followers/friends who also participated in the earthquake information dissemination by posting/reposting Sina-Weibo ( $WU.sCount-P$  and  $WU.repostsCount-P$ ), which are counted by the following equations:

$$WU.frCount(wu) = |\{wu.friendSet\}| \quad (1)$$

$$WU.frCount(wu) = |\{wu.followerSet\}| \quad (2)$$

$$WU.frCount - p(wu) = |\{u \mid u \in wu.friendSet \text{ and } u \in WUS\}| \quad (3)$$

$$WU.frCount - p(wu) = |\{u \mid u \in wu.followerSet \text{ and } u \in WUS\}|. \quad (4)$$

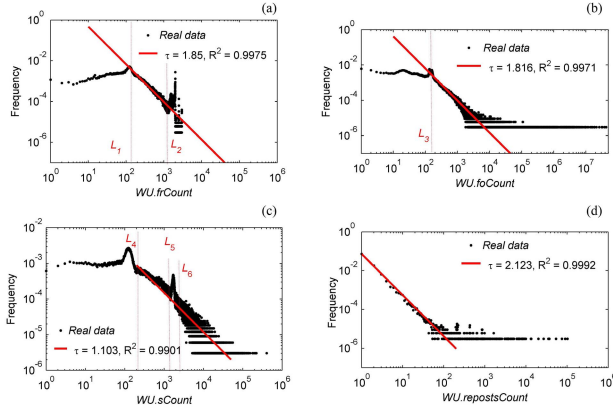


Fig. 5. Distributions of four individual attributes of participant users after the Yi'liang earthquake. (a) Number of each participant's friends, and the fitting interval is located  $[130, 1300]$  between  $L_1$  and  $L_2$ . (b) Number of each participant's followers, and the fitting interval is located  $[130, +\infty)$  on the right side of  $L_3$ . (c) and (d) Number of each participant's friends who also participated in the earthquake information dissemination by posting/reposting Sina-Weibo messages. In addition, " $\tau$ " is the slope of fitting line, " $R^2$ " is the coefficient of determination of the fitting procedure.

The fitting function of red fitting lines is the general power model—frequency( $X = x$ ) =  $c \cdot x^{-\tau}$  with 95% confidence bounds, where  $c, \tau \in \mathbb{R}^+$ , and  $x \in \mathbb{N}^+$ .

Fig. 5 shows the distributions of four individual attributes of participant users. Sudden spikes in Fig. 5(a)–(c) are very obvious. However, we thought that these phenomena have nothing to do with social networks, but are led by the constraints of users privileges on Sina-Weibo. Just with the upper limit of  $WU.frCount$ , for example, there are three levels: 2000, 2500, and 3000 for nonmembers, ordinary members, and VIP members, respectively (see <http://vip.weibo.com/privilege>).

Another common feature found in Fig. 5 is segment characteristics. This is true in Fig. 5(a) and (b), where the sudden changes near 130 verify Dunbar's number. That is, human communities are much larger than those of other primates and hence require more time to be devoted to social maintenance activities. However, there is an upper limit on the amount of time that can be dedicated to social demands, so this sets an upper limit on social group size [41], [42]. It is known that humans have the cognitive capacity to maintain about 150 stable social relationships. With the advent of different types of super social networking services developing one after another, people have once again picked up the topic for discussion [43], [44]. Many researchers have investigated how tools, such as Facebook and Twitter, have changed our capacity to handle social connections using the empirical study [45]–[47]. Here, Fig. 5(a) and (b) also shows the shadow of Dunbar's number on Sina-Weibo.

In addition, compared with the distribution of broadly defined "friends/followers" which are influenced by the Sina-Weibo user privilege in Fig. 5(a) and (b), the number of each participant's "actual friends/followers" in Fig. 5(c) and (d) shows the typical power-law distribution; from this, we could know that the unevenness is a universal phenomenon of social networks, even when there are some limiting conditions in social network services, such as the constraint of users' privileges on Sina-Weibo as described earlier.

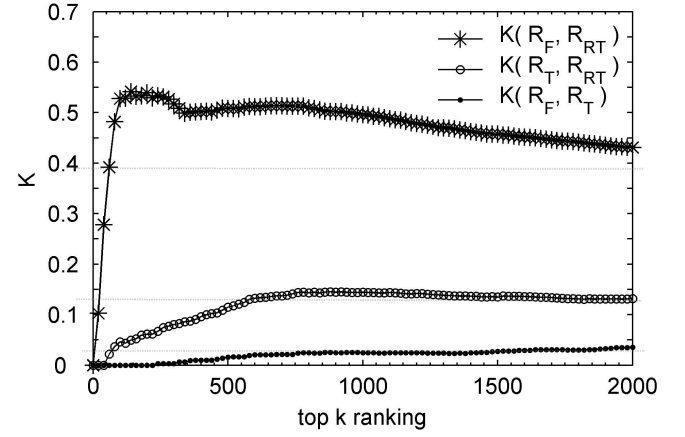


Fig. 6. Comparison between three types of user rankings for all users by  $R_F$ ,  $R_{RT}$ , and  $R_T$ .

2) *User Influence*: The number of relevant tweets is the most immediate indication of the popularity of posts on the Weibo space. The number of being reposted can offer us an intuitive view of the original poster's influence on public opinion. For the popularity, the engagement/productivity, and the influence [5], [40], [48], three types of user rankings are made by the following  $R_T$ ,  $R_F$ , and  $R_{RT}$

$$R_F(wu|wu \in WUS) = |wu.friendSet| \quad (5)$$

$$R_F(wu|wu \in WUS) = |\{T|T \in TS \wedge T.user = wu\}| \quad (6)$$

$$R_{RT}(wu|wu \in WUS) = \sum_{T \in TS} (T.repostCount|T.user = wu). \quad (7)$$

For the top 20 users of each ranking, the comparison shows that entertainers belong to the most popular group and most users who post tweets actively come from the grassroots, but nearly all users whose tweets regularly receive widespread reposting are news/charity organizations or entertainers. The rest of this subsection quantifies the correlations between the three rankings by the generalized- $\tau$  model [49]:

$$K_{\tau}^{(0)}(R_1, R_2) = \sum_{\{r_1, r_2\} \in (R_1 \cup R_2)} \bar{K}_{r_1, r_2}(R_1, R_2) \quad (8)$$

where  $R_1$  and  $R_2$  are two ordered rankings with the equal length,  $k$ ; and  $\bar{K}_{r_1, r_2}(R_1, R_2) = 1$ , if 1)  $r_1, r_2 \notin R_1 \cap R_2$  and  $r_1$  is only in one ranking and  $r_2$  is in the other ranking; 2)  $r_1$  is ranked higher than  $r_2$  and only  $r_2$  appears in the other ranking; or 3)  $r_1$  and  $r_2$  are in both rankings in the opposite order, otherwise,  $\bar{K}_{r_1, r_2}(R_1, R_2) = 0$ . In particular,  $\bar{K}_{r_1, r_2}(R_1, R_2) = \bar{K}_{r_2, r_1}(R_1, R_2)$ . In addition, the normalized distance- $K$  is used, computed as follows [50]:

$$K(R_1, R_2) = 1 - \frac{K_{\tau}^{(0)}(R_1, R_2)}{k^2}, \quad \text{where } k = |R_1| = |R_2|. \quad (9)$$

The range of  $K(R_1, R_2)$  is from 0 to 1.  $K(R_1, R_2) = 0$  means complete disagreement, and  $K(R_1, R_2) = 1$  means complete agreement.

Fig. 6 shows that with the increase of  $k$ , the distance between two arbitrary rankings will get to a relatively stable

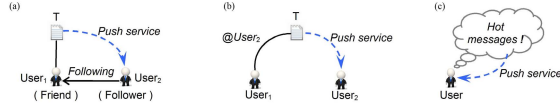


Fig. 7. Three main tweet push services on Sina-Weibo. (a) For any two users,  $User_1$  and  $User_2$ , if  $User_2$  follows  $User_1$ , then  $User_1$  is the friend of  $User_2$  and  $User_2$  is the follower of  $User_1$ , in addition, tweets posted by  $User_1$  would be shown for  $User_2$  in real time. (b) All tweets beginning with the string of “@ $User_2$ ” will be shown for  $User_2$ . (c) Every user can read “hot messages” that Sina-Weibo picked out from all tweets of a certain period by their popularity.

level, but three stable levels vary widely. 1) The top association between  $R_F$  and  $R_{RT}$  allow us to infer that a huge fan base may indeed be transformed into a real influence on the public. 2) The weak but stable association between  $R_T$  and  $R_{RT}$  let us deduce that active participation may indeed help to enhance user influence, but it is very limited. 3) There is little or no association between  $R_T$  and  $R_F$ , so it can be inferred that active grassroots participators seldom intersect with celebrities. Therefore, it can be seen that Weibo is a place shared by the general public and celebrities, which makes Weibo an excellent example of grassroots media. The seldom intersection makes the metamorphosis from grassroots activist to celebrity difficult, but it is a truth that intimate contact can make overnight success possible, though we have not found the hidden principles among this [51].

### B. User Relationship Structure Measurement

From a sociological perspective, friendship is a trust relationship built by repeated games, and now the quality of online social relationships has been receiving increased attention [30], [31], [52]. In our earthquake discussion case, Fig. 7 lists the three main information push services in Sina-Weibo. This section focuses on the following questions. Who actually repost tweets that the original user has posted? When a user reposts, does he/she repost tweets that he/she is interested in, or just those tweets posted by his/her friends? Also for users who get reposted, do the reposting all take place just among their followers? Are there any differences in information diffusion patterns before and after the earthquake?

1) *Structure of Weibo Information Flow*: For the five types of  $RR$  in Fig. 3, the butterfly shaped cartoon structure in Fig. 8 analyzed their share in the total amount of  $RR$  before and after earthquake, respectively, computed as follows:

$$Ratio(RRS_i) = \frac{|RRS_i|}{|RRS|} \times 100\% \quad (10)$$

where  $RRS_i = \{RR | RR \in RRS \wedge RR.type = i\}$ ,  $i \in \{I, II, III, IV, V\}$ .

For the WIF structure before the Ya'an earthquake, Fig. 8(a) shows as follows. First, the  $RR_{type=III}$  from  $WU_0$  to  $WU_3$  shares the largest percentage of all  $RR$ s, 43.04%, which benefits from the push service in Fig. 7(a). Second, there is no declared relationship between  $WU_0$  and  $WU_4$ , but the share of  $RR_{type=IV}$  between them reaches up to 37.23%, which may be largely due to the push service in Fig. 7(c). Also, it can be seen that compared with the user popularity, the content of

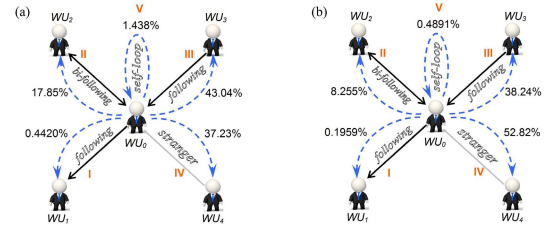


Fig. 8. Structure of Sina-Weibo information spread before and after the earthquake. Notes: five parts marked with I~IV correspond to the five types of  $RR$ s in Fig. 3, and the percentages are their shares in the total amount of all  $RR$ s. Structure of Sina-Weibo information spread during (a) week before the Ya'an earthquake and (b) seven months after Yi'liang earthquake.

tweets is equally important for their great popularity. Third, the strongest declared friendship exists between  $WU_0$  and  $WU_2$ , and the share of  $RR_{type=II}$  between them is 17.85%, which can be considered the mirror of the contacting pattern among friends in real life. Fourth, the share of  $RR_{type=V}$  is 1.438%. Unfortunately, we found no logical explanation for these behaviors. Fifth, the  $RR_{type=I}$  from  $WU_0$  to  $WU_1$  shares the smallest percentage of 0.4420%, less than 1%. Most users did not pay attention to tweets posted by their followers at all. It can be noticed that there is not an information push service from followers to friends, but in the real world, there is also a limited number of efficient ways to attract celebrities. So, this is what we can see in the hedged real world just through the lens of online media.

Fig. 8(b) shows the WIF structure after the Yi'liang earthquake, and in Table II, we have listed the comparison result of these two butterfly maps. We can see the proportion of  $RR_{type=IV}$  from  $WU_0$  to  $WU_4$  rises remarkably, the growth is more than half.

In this subsection, the following three conclusions can be drawn. First, “3.1 Analysis of user characteristics” has already shown that very few Weibo-big-Vs are able to attract wide attention; in other words, an overwhelming majority of Weibo users play the role of followers. From the dramatic contrast between proportions of  $RR_{type=I}$  and  $RR_{type=III}$ , Sina-Weibo seems like a natural platform to worship idols. Second, given the dramatic contrast among the proportions of  $RR_{type=II}$  and  $RR_{type=III}$  or  $RR_{type=IV}$ , Weibo users pay more attention to receiving interesting information than making friends. Maybe this is what makes Weibo different from other online social platforms. Third, followers are indeed the major driving force for the Weibo information spread, but at the same time, the reports from strangers are equally crucial. Particularly in some public emergencies, strangers are likely to be in the vanguard of discussing and distributing news in real time. So, from the point of view of information distribution, in addition to the followers' count of a user, it is also important to examine whether the tweet is selected as a hot message used to push to all other users (see <http://hot.weibo.com/>).

2) *Distributions of Individual Being Reposted Times*: This subsection presents the distributions of individual reposting times in each type of  $RRS$  for the WIF structure after the Yi'liang earthquake [Fig. 8(b)]. In Fig. 9(f), the statistic sample



TABLE II  
SHARE OF EACH TYPE OF RRS IN WIF BEFORE AND AFTER EARTHQUAKE

Social context of the DataSet	Multi-Column				
	Type=I	Type=II	Type=III	Type=IV	Type=V
Before Ya'an Earthquake	0.4420%	17.85%	43.04%	37.23%	1.438%
After Yi'liang Earthquake	0.1959%	8.255%	38.05%	52.82%	0.4891%

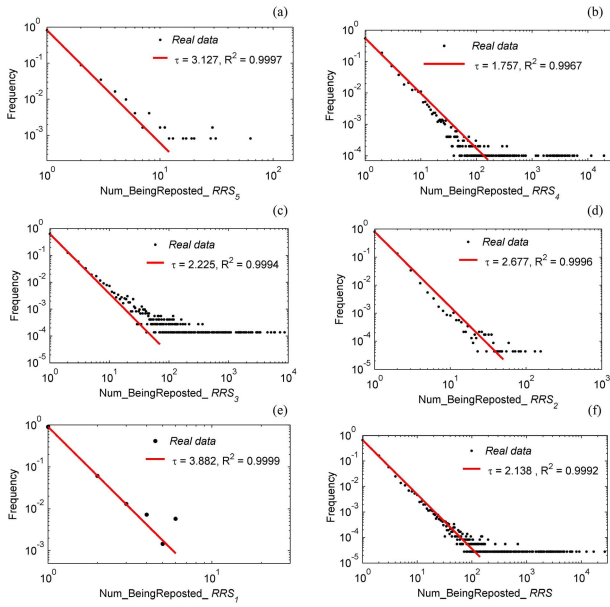


Fig. 9. Distributions of the number of being reposted of each *WU* in *RRS* grouped by *RR.type*. (a) *RRS* of Weibo users' own reposts. (b) *RRS* of reposts from strangers. (c) *RRS* of reposts from followers. (d) *RRS* of reposts from bifriends (mutually following each other). (e) *RRS* of reposts from friends. (f) *RRS* of all reposts.

set is the sum of all *RRs*. If there is a point  $(k, p_k)$  in the scatter plot, it means that those users whose tweets have been reposted  $k$  times by others makeup  $p_k$  percentage of participating users total, and there is a following mapping between  $k$  and  $p_k$ :

$$p_k = \frac{\sum_{wu \in WUS} count(wu)}{|\{wu | \exists RR (RR \in RRS \wedge RR.st.user = wu)\}|} \quad (11)$$

where if the following equation is true, then  $count(wu) = 1$ , otherwise  $count(wu) = 0$ :

$$|\{RR | RR \in RRS \wedge RR.st.user = wu\}| = k. \quad (12)$$

The five statistic sample sets in Fig. 9(a) and (e) are  $RRS_Y$ ,  $RRS_{IV}$ ,  $RRS_{III}$ ,  $RRS_{II}$ , and  $RRS_I$ , respectively. So, the point  $(k, p_k)$  in these five scatter plots means that there is a following mapping between  $k$  and  $p_k$ :

$$p_k = \frac{\sum_{wu \in WUS} count(wu)}{|\{wu | \exists RR (RR \in RRS_i \wedge RR.st.user = wu)\}|} \quad (13)$$

where  $i \in \{I, II, III, IV, V\}$ , and if the following equation is true, then  $count(wu) = 1$ , otherwise  $count(wu) = 0$ :

$$|\{RR | RR \in RRS_i \wedge RR.st.user = wu\}| = k. \quad (14)$$

All six distributions in Fig. 9 follow a power-law distribution. This is consistent with the distributions of the number of being cited as well as citing others in an empirical study

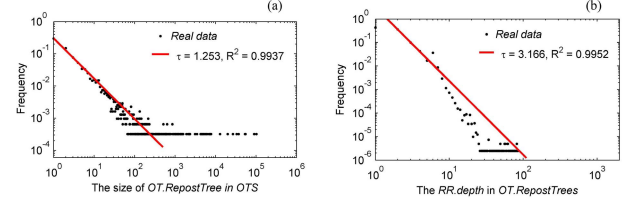


Fig. 10. Distribution of *TRT.size* and *RR.depth* in the *WIF* related to the Yi'liang earthquake. (a) *TRT.size* distribution. (b) *RR.depth* distribution.

of HFS<sup>6</sup>. However, the three similar power-law slope values ( $\tau = 1.68, 1.75$ , and  $1.84$ ) in the HFS study are equivalent to one distribution ( $\tau = 1.757$ ) of  $RR_{type=IV}$ , far below those ( $\tau = 3.127, 2.225, 2.677, 3.882$ , and  $2.138$ ) other five distributions. The nature underlying power-law distribution is uneven, and the higher a slope value is, the more severe the imbalance gets. Corresponding to the situation here, it means that these users participating  $RR_{type=IV}$  maintain the same interaction pattern with users of ordinary online social forums. However, there might be a serious structural imbalance in the interactions of  $RR_{type=I}$ ,  $RR_{type=II}$ ,  $RR_{type=III}$ , and  $RR_{type=V}$ , which may be the reason why Weibo is so different, churning out a string of “Weibo-big-Vs”-Weibo users with mass followings and whose identities have been verified by Sina. These features have to do with social role partitioning [Fig. 3(a)] and the role-based push services (Fig. 7) on Sina-Weibo.

### C. Topology Measurement of Real Interaction Networks

Research by Huberman *et al.* [28] found that the driver of Twitter usage is a sparse and hidden network of connections underlying the “declared” set of friends and followers [53], that is to say in terms of the network density, there is a gap between the real interaction network and the declared user network. Additionally, some other researchers have also found that social interaction existed among various types of users, far more than among acquaintances [54]. From the perspective of network topology, beyond the network density, are there any differences in other topological properties? In addition, what are the key factors behind user influence? To answer these questions, first, a method is introduced to extract the hidden interaction network from the friends and followers network on the basis of the *WIF* model. Afterward, we present measurement results and provide the corresponding explanation.

The *TRT* structure is the key to extracting real interaction networks. For 3096 *TRTS* in the *WIF* after the Yi'liang earthquake, Fig. 10(a) shows their size distribution where the largest *TRT* consists of 102 526 retweets, and Fig. 10(b) shows

the *depth* distribution of all *RRs*, where the maximum value is 85.

1) *Extraction Process of Two Types of Real Interaction Networks*: According to the definition of WIF presented in Section II-B, the symbolic representation of the WIF instance in Fig. 11 is listed as follows:

$$WIF = \langle WUS, TS, RRS, TRTS \rangle$$

where  $WUS = \{wu_2, wu_3, wu_4, wu_5, wu_6, wu_7, wu_8, wu_9, wu_{10}, wu_{11}, wu_{12}\}$ .  $TS = \{ot_1, ot_2, ot_3, ot_4, ot_5, ot_6, ot_7, ot_8, ot_9, ot_{10}, ot_{11}, ot_{12}, ot_{13}\}$ , and  $ot_1 = \langle 0, wu_5, 0, -, -, null \rangle$ ,  $ot_2 = \langle 0, wu_9, 3, -, -, null \rangle$ ,  $ot_3 = \langle 0, wu_1, 10, -, -, null \rangle$ ,  $rt_1 = \langle 1, wu_{10}, 1, -, -, ot_2 \rangle$ ,  $rt_2 = \langle 1, wu_4, 0, -, -, ot_2 \rangle$ ,  $rt_3 = \langle 1, wu_{11}, 0, -, -, ot_2 \rangle$ ,  $rt_4 = \langle 1, wu_2, 2, -, -, ot_3 \rangle$ ,  $rt_5 = \langle 1, wu_2, 5, -, -, ot_3 \rangle$ ,  $rt_6 = \langle 1, wu_{12}, 0, -, -, ot_3 \rangle$ ,  $rt_7 = \langle 1, wu_3, 1, -, -, ot_3 \rangle$ ,  $rt_8 = \langle 1, wu_6, 1, -, -, ot_3 \rangle$ ,  $rt_9 = \langle 1, wu_7, 2, -, -, ot_3 \rangle$ ,  $rt_{10} = \langle 1, wu_7, 0, -, -, ot_3 \rangle$ ,  $rt_{11} = \langle 1, wu_6, 0, -, -, ot_3 \rangle$ ,  $rt_{12} = \langle 1, wu_8, 0, -, -, ot_3 \rangle$ ,  $rt_{13} = \langle 1, wu_3, 0, -, -, ot_3 \rangle$ .  $RRS = \{rr_1, rr_2, rr_3, rr_4, rr_5, rr_6, rr_7, rr_8, rr_9, rr_{10}, rr_{11}, rr_{12}, rr_{13}\}$ , and  $rr_1 = \langle ot_2, rt_1, 1, 1 \rangle$ ,  $rr_2 = \langle ot_2, rt_2, 1, 4 \rangle$ ,  $rr_3 = \langle rt_1, rt_3, 2, 4 \rangle$ ,  $rr_4 = \langle ot_3, rt_4, 1, 3 \rangle$ ,  $rr_5 = \langle ot_3, rt_5, 1, 3 \rangle$ ,  $rr_6 = \langle ot_3, rt_6, 1, 4 \rangle$ ,  $rr_7 = \langle rt_4, rt_7, 2, 3 \rangle$ ,  $rr_8 = \langle rt_5, rt_8, 2, 2 \rangle$ ,  $rr_9 = \langle rt_5, rt_9, 2, 4 \rangle$ ,  $rr_{10} = \langle rt_7, rt_{10}, 3, 4 \rangle$ ,  $rr_{11} = \langle rt_8, rt_{11}, 3, 5 \rangle$ ,  $rr_{12} = \langle rt_9, rt_{12}, 3, 4 \rangle$ ,  $rr_{13} = \langle rt_9, rt_{13}, 3, 4 \rangle$ .  $TRTS = \{trt_1, trt_2, trt_3\}$ , and  $trt_1 = \langle ot_1, null, null, 1 \rangle$ ,  $trt_2 = \langle ot_2, \{rt_1, rt_2, rt_3\}, \{rr_1, rr_2, rr_3\}, 4 \rangle$ ,  $trt_3 = \langle ot_3, \{rt_4, rt_5, rt_6, rt_7, rt_8, rt_9, rt_{10}, rt_{11}, rt_{12}, rt_{13}\}, \{rr_4, rr_5, rr_6, rr_7, rr_8, rr_9, rr_{10}, rr_{11}, rr_{12}, rr_{13}\}, 11 \rangle$ .

The “declared” online social network is still the common empirical object among existing empirical studies, where nodes represent users, and the edge between nodes indicates the declared social relationship [6]–[8], [39], [48]. Given the prevalence of the Internet water army, this paper is only interested in how participants collaborated with each other and their relationship types. In addition, we found

$$\frac{|\{RR | RR \in RRS \wedge RR.depth = 1 \wedge RR.rt.repostCount = 0\}|}{|RRS|} = 38.85\%$$

that is, more than a third of retweets in TRTS link only to the original tweet without any citations relating to other retweets, such as  $rt_2$  and  $rt_6$ , in Fig. 11(a). We denoted these types of retweets as casual nodes and the corresponding participants as casual participants. Although casual nodes help spread information (the total number of reposted tweets is an important factor for tweet rank), those nodes did not contribute to the actual collaboration activities. Therefore, we excluded casual nodes and analyzed the remaining repost behavior, which involved a total of 249237 retweets.

Fig. 12(a) shows one type of real interaction network extracted from the WIF instance in Fig. 11, which was named the friendship-based reposting cooperation network (FRCN). The cartoon figures represent distinct participants, and directed edges indicate the type of social relationships between them [corresponding to the black directed edges Fig. 3(a)]. The extraction process needs to visit all directed edges in Fig. 11(b), retaining only the edges and the participants linked by edges that match the following conditions.

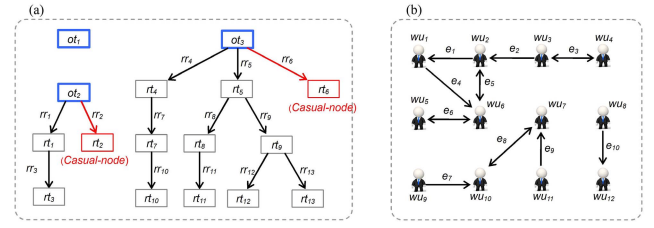


Fig. 11. Example of the WIF included 16 Weibo messages and 12 Weibo users. (a) Structure of the Tweet Repost Tree Set, where each node is corresponding to a unique Weibo message, and directed edges between pairs of nodes indicate the presence of posting citations between them. (b) Structure of the following-based Weibo user declared social network, where each cartoon guy is corresponding to a unique Weibo user, and directed edges between pairs of guys indicate the presence of the following relationship between them.

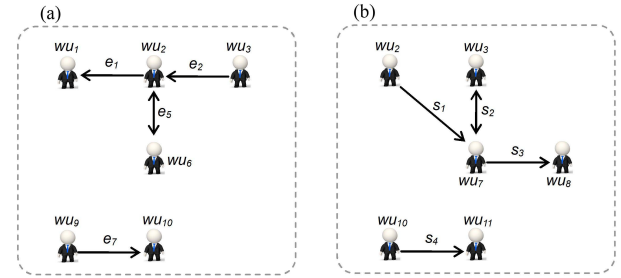


Fig. 12. Two types of Weibo message spread participant network. (a) FRCN the FRCN. (b) SRCN the stranger reposting cooperation network.

That is, if there is a pair of users ( $wu_i, wu_j$ ) linked by edges in FRCN, then in Fig. 11(a) there is at least one *RR* that makes the following formula true:

$$\exists RR (RR \in RRS \wedge RR.st.user = wu_i \wedge RR.rt.user = wu_j). \quad (15)$$

Another type of real interaction network is shown in Fig. 12(b); named stranger reposting cooperation network (SRCN) where the cartoon figures represent distinct participants, and directed edges indicate the information flow between them (corresponding the blue directed edges in Fig. 3). The extraction process needs to visit all *RRs*, where type = IV in Fig. 11(b). If there is at least one *RR* matching the following condition, then we add the corresponding pair of users ( $wu_k, wu_l$ ) and one directed edge from  $wu_k$  to  $wu_l$  into SRCN:

$$\exists RR (RR \in RRS \wedge RR.type = 4 \wedge RR.ST.user = wu_k \wedge RR.RT.user = wu_l).$$

Finally, on the basis of the WIF after Yi'liang earthquake, the FRCN consisted of 128865 nodes and 159379 edges and the SRCN consisted of 85978 nodes and 104410 edges. Therefore, it can be seen again that Sina-Weibo is an excellent synthesis of the traditional society of acquaintances and the strangers' one.

2) *Topology Measurement of FRCN and SRCN*: Table III lists a comparison of topological properties between FRCN, SRCN, and HFS7; some corresponding distributions are presented in Fig. 13.



TABLE III  
TOPOLOGICAL PROPERTIES COMPARISON OF THE HFS, THE FRCN, AND THE SRCN

Networks	$N$	$E$	$N_C$	$N_G$	$\rho$	$\tau_{in}$	$\tau_{out}$	$C$	$avg.D$	$S$	$d$
HFS <sup>7</sup>	20 813	29 798	2 821	11 556(55.5%)	$10^{-4}$	2.1	2.4	0.027	2.650	8.679	28
FRCN	128 865	159 379	5 422	112 617(87.4%)	$10^{-5}$	2.813	2.318	0.003	1.998	10.54	47
SRCN	85 978	104 410	104 410	6 241	$10^{-5}$	2.848	1.887	$10^{-4}$	2.401	1.745	9

$N$ : number of nodes;  $E$ : number of edges;  $N_C$ : number of components;  $N_G$ : number of nodes in the giant component;  $\rho$ : network density;  $\tau_{in}$ : power of in-degree distribution;  $\tau_{out}$ : power of out-degree distribution;  $C$ : average clustering coefficient;  $avg.D$ : average degree;  $S$ : average shortest path length;  $d$ : network diameter.

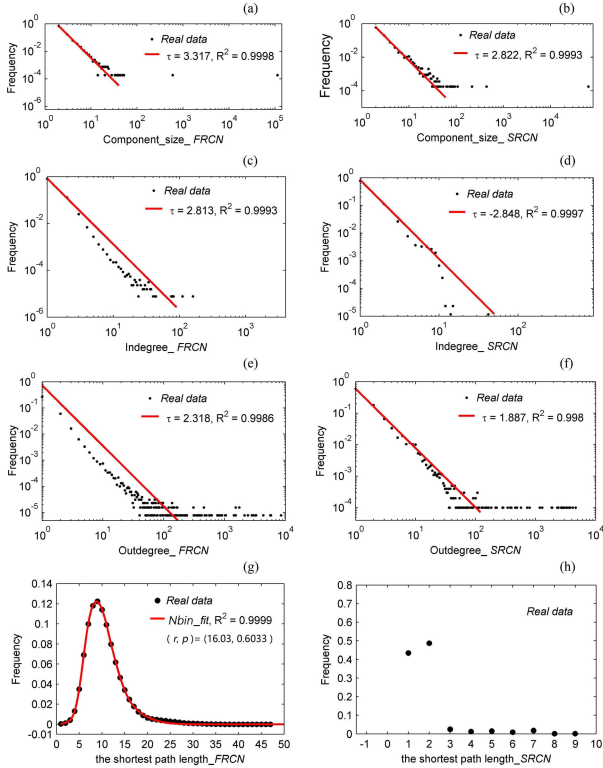


Fig. 13. Distributions of three topological properties in FRCN and SRCN. (a) Number of nodes in each component of FRCN. (b) Number of nodes in each component of SRCN. (c) In-degree of each node in FRCN. (d) In-degree of each node in SRCN. (e) Out-degree of each node in FRCN. (f) Out-degree of each node in SRCN. (g) Length of each shortest path in FRCN. (h) Length of each shortest path in SRCN.

The values of  $N_C$  and  $N_G$  in Table III show that there is a largest weekly connected component (giant component) both in FRCN and SRCN, and Figs. 13(a) and 14(b) show that the component size distributions of FRCN and SRCN both follow the power-law trait. Connectivity is an important part of the social network structure, and our analysis results indicate that doing things in groups is a strong human instinct; that is, everything we do or say tends to ripple through our network and have an impact on our friends. For the context of our study, if interest and friendship created the emergence of super components in the HFS groups and FRCN, then what is it that made the emergence of super components in SRCN? Though this question is very interesting, it goes beyond the scope of this paper.

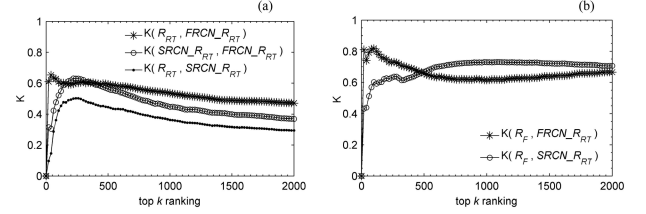


Fig. 14. Comparison of user rankings between FRCN and SRCN. (a) Comparison between  $R_{RT}$ ,  $SRCN_{R_{RT}}$ , and  $FRCN_{R_{RT}}$ . (b) Comparison between  $R_F$ ,  $SRCN_{R_{RT}}$ , and  $FRCN_{R_{RT}}$ .

The comparison of  $\rho$  in Table III shows that values of the network densities of FRCN and SRCN are both much less than the HFS groups; the interaction groups in Sina-Weibo are looser and more disorganized than traditional online social forums, such as HFS communities. This allows Weibo users to group together quickly and leave quickly.

All of the degree distributions in Fig. 13(c) and (f) are the power-law type, meaning that FRCN and SRCN both are scale-free networks. This also means that there are users whose tweets always receive the overwhelming public response (i.e., huge hubs), but the appeal of most of others is severely limited. For the scale-free property in a social network, the research team of Albert-László Barabás holds that it is a consequence of a decision-based queuing process: when individuals execute tasks based on some perceived priority, there will be heavy tail phenomenon [55]. We do not rule out the possibility, because the number being reposted is indeed an important factor for a tweet in a real-time-based push service shown in Fig. 7(c). However, one thing is certain: without the strong appeal of hub users in either FRCN or SRCN, few messages could get widespread.

The comparison of  $C$ ,  $avg.D$ ,  $S$ , and  $d$  in Table III shows that FRCN and SRCN both are small-world networks. The  $avg.D$  values of FRCN and SRCN are both below the  $avg.D$  of HFS, and their  $C$  values are at least an order of magnitude lower than HFS, indicating that most users are not friends of one another. However, the values of  $S$  and  $d$  in Table III and the result in Fig. 13(g) and (h) both shows that the  $avg.D$  has not affected the social distance, i.e., Weibo information can always flow from one user to another through a small number of hops. With the SRCN, for example, each user has less than three adjacent users on average, but the typical distance between randomly chosen users still remains fewer than nine hops. Another finding about FRCN and SRCN is that

the shortest paths of FRCN behave according to the negative binomial distribution.

Finally, on the basis of analysis in Fig. 6, the rest of this section performed further analysis on how much the influential group may converge in FRCN and SRCN with the generalized *Kendall*– $\tau$  model again. FRCN and SRCN are two real interaction networks emerging in entirely different ways, so we wondered in particular whether the appeal of these users on the top of  $R_F$  and  $R_{RT}$  maintain the same powerful influence both in FRCN and SRCN. The two formulas in the following are for ranking users of FRCN and SRCN, respectively. And the results are shown in Fig. 14

$$FRCN_{RT}(wu) = |\{RR | RR \in RRS \wedge \neg(RR.type = 4) \wedge RR.st.user = wu\}| \quad (16)$$

$$SRCN_{RT}(wu) = |\{RR | RR \in RRS \wedge RR.type = 4 \wedge RR.st.user = wu\}|. \quad (17)$$

The total number of times a tweet is reposted is the most concrete indication of user influence. The comparison between the two curves of  $K(R_{RT}, FRCN_{RT})$  and  $K(R_{RT}, SRCN_{RT})$  in Fig. 14. A indicates that the influential users in FRCN cater to popular taste better than those in SRCN. The curve of  $K(SRCN_{RT}, FRCN_{RT})$  in Fig. 14(a) shows that an influential group does converge in FRCN and SRCN, but with limited overlap especially so for the large-scale and widespread reposting. In addition, the two curves in Fig. 14(b) have a similar level, it can be observed that the two influential groups in FRCN and SRCN have at least one thing in common: both have a massive following of fans. Therefore, the determinant of a tweet gets large-scale notice would likely be related to the number of followers the publisher has.

#### IV. CONCLUSION

Weibo has been ubiquitously integrated into people's everyday lives in China. Both Sina-Weibo and Tencent-Weibo had more than five million users in early May 2013, and their open platforms (including the data API service) have been improved constantly. Although the very existence of user records raises huge concerns of privacy, these user records also create a historic opportunity, that is offering for the first time unbiased data of unparalleled detail on the behavior of not one, but millions of individuals.

In this paper, we performed a comprehensive analysis of Weibo information diffusion during earthquakes. We found that symbolic representation applied to the WIF model is indeed a feasible choice for the empirical study of human behavior based on online social media data sets. In retrospect, our primary inspiration came from the description mechanism of concepts and relationships in ontology theory. The main feature of this idea is that it can give a formal expression for the data structure and the analysis process (such as the extraction of FRCN and SRCN).

However, the structure of social networks is only a starting point. When people talk about the "connectedness" of a social network, in general, they are really talking about two related issues. One is who is linked to whom; and the other is the fact that each individual's actions have implicit consequences

for the outcomes of everyone in the system [56]. In fact, Fig. 8 has given us some intuition that there is probably a serious structural imbalance between the "declared" relationship network and the real interaction network [57]–[59]. In addition, to measure public perceptions in emergencies, many researchers have worked extensively on the evolution of public opinion during information dissemination based on Twitter [60], which allows for many interesting directions for future work.

#### REFERENCES

- [1] Defense Advanced Research Projects Agency. (2010). *DARPA Network Challenge Project Report*. [Online]. Available: <http://www.eecs.harvard.edu/cs286r/courses/fall10/papers/ProjectReport.pdf>
- [2] G. Pickard *et al.*, "Time-critical social mobilization," *Science*, vol. 334, no. 6055, pp. 509–512, Oct. 2011.
- [3] M. Kitsak *et al.*, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, no. 11, pp. 888–893, Aug. 2010.
- [4] E. Bakshy, J. M. Hofman, D. J. Watts, and W. A. Mason, "Everyone's an influencer: Quantifying influence on twitter," in *Proc. WSDM*, Hong Kong, 2011, pp. 65–74.
- [5] K.-W. Fu and M. Chau, "Reality check for the Chinese microblog space: A random sampling approach," *PLoS ONE*, vol. 8, no. 3, p. e58356, Mar. 2013.
- [6] P. Fan, H. Wang, Z. Jiang, and P. Li, "Measurement of microblogging network," (in Chinese), *J. Comput. Res. Develop.*, vol. 49, no. 4, pp. 691–699, 2012.
- [7] K. Lei, K. Zhang, and K. Xu, "Understanding Sina Weibo online social network: A community approach," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3114–3119.
- [8] Z. Zhan, Y. Chen, and Y. Fu, "Analyzing user relationships in Weibo networks: A Bayesian network approach," in *Proc. CloudCom-Asia*, Fuzhou, China, 2014, pp. 502–508.
- [9] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Phys. Rep.*, vol. 519, no. 1, pp. 1–49, 2012.
- [10] F. Fu, L. Liu, and L. Wang, "Empirical analysis of online social networks in the age of Web 2.0," *Phys. A, Statist. Mech. Appl.*, vol. 387, nos. 2–3, pp. 675–684, Jan. 2008.
- [11] L. L. Yu, S. Asur, and B. A. Huberman, "Artificial inflation: The real story of trends and trend-setters in Sina Weibo," in *Proc. Int. Conf. Privacy, Secur., Risk Trust (PASSAT), Int. Conf. Social Comput. (SocialCom)*, Sep. 2012, pp. 514–519.
- [12] M. T. Bastos, R. Travitzki, and R. Raimundo. (2012). *Tweeting political dissent: Retweets as pamphlets in FreeIran, FreeVenezuela, Jan25, SpanishRevolution and OccupyWallSt. Integrated Performance Primitives*. [Online]. Available: [http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/BASTOS-TRAVITZKI-RAIMUNDO\\_-\\_Tweeting\\_Political\\_Dissent.pdf](http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/BASTOS-TRAVITZKI-RAIMUNDO_-_Tweeting_Political_Dissent.pdf)
- [13] L. A. Overbey, C. Paribello, and T. Jackson, "Identifying influential twitter users in the 2011 egyptian revolution," in *Proc. SBP*, 2013, pp. 377–385.
- [14] K. Lerman, R. Ghosh, and T. Surachawala. (2013). "Social contagion: An empirical study of information spread on digg and twitter follower graphs." [Online]. Available: <http://arxiv.org/abs/1202.3162>
- [15] G. V. Steeg, R. Ghosh, and K. Lerman. (2011). "What stops social epidemics?" [Online]. Available: <http://arxiv.org/abs/1102.1985>
- [16] A. Guille, H. Hacid, and C. Favre. (2013). "Predicting the temporal dynamics of information diffusion in social networks." [Online]. Available: [http://arxiv.org/abs/1302.5235?utm\\_source=twitterfeed&utm\\_medium=twitter](http://arxiv.org/abs/1302.5235?utm_source=twitterfeed&utm_medium=twitter)
- [17] J. P. Bagrow, D. Wang, and A.-László Barabási, "Collective response of human populations to large-scale emergencies," *PLoS ONE*, vol. 6, no. 3, p. e17680, 2011.
- [18] R. M. Allen, "Transforming earthquake detection?" *Science*, vol. 335, no. 6066, pp. 297–298, Jan. 2012.
- [19] F.-Y. Wang *et al.*, "A study of the human flesh search engine: Crowd-powered expansion of online knowledge," *Computer*, vol. 43, no. 8, pp. 45–53, Aug. 2010.
- [20] Q. Zhang *et al.*, "Understanding crowd-powered search groups: A social network perspective," *PLoS ONE*, vol. 7, no. 6, pp. 1–16, Jun. 2012.

- [21] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time event detection by social sensors," in *Proc. WWW*, Raleigh, NC, USA, 2010, pp. 851–860.
- [22] B. Robinson, R. Power, and M. Cameron, "A sensitive twitter earthquake detector," in *Proc. Companion WWW*, Rio de Janeiro, Brazil, 2013, pp. 999–1002.
- [23] A. Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski, "Earthquake: Twitter as a distributed sensor system," *Trans. GIS*, vol. 17, no. 1, pp. 124–147, Feb. 2013.
- [24] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 919–931, Apr. 2013.
- [25] M. Moriya and M. Ryoke, "Information balance between transmitters and receivers based on the twitter after great East Japan earthquake," *Int. J. Knowl. Syst. Sci.*, vol. 4, no. 2, pp. 66–76, Apr. 2013.
- [26] J. Y. Jung and M. Moro, "Multi-level functionality of social media in the aftermath of the great East Japan earthquake," *Disasters*, vol. 38, no. 2, pp. S123–S143, Jul. 2014.
- [27] F. Toriumi, T. Sakaki, K. Kazama, S. Kurihara, I. Noda, and K. Shinoda, "Information sharing on Twitter during the 2011 catastrophic earthquake," in *Proc. WWW Companion*, Rio de Janeiro, Brazil, 2013, pp. 1025–1028.
- [28] B. Huberman, D. M. Romero, and F. Wu. (Dec. 2008). Social networks that matter: Twitter under the microscope. First Monday, [S.l.], accessed: Jan. 2, 2018. [Online]. Available: <http://firstmonday.org/ojs/index.php/fm/article/view/2317/2063>, doi: <http://dx.doi.org/10.5210/fm.v14i1.2317>.
- [29] M. Tavakolifard, K. C. Almeroth, and J. A. Gulla, "Does social contact matter?: Modelling the hidden Web of trust underlying Twitter," in *Proc. WWW Companion*, Rio de Janeiro, Brazil, 2013, pp. 981–988.
- [30] S. K. Lam and J. Riedl, "Are our online 'Friends' Really Friends?" *Comput.*, vol. 45, no. 1, pp. 91–93, Jan. 2012.
- [31] S. Wu, J. M. Hofman, and W. A. Mason, "Who says what to whom on twitter," in *Proc. Int. Conf. World Wide Web (WWW)*, Hyderabad, India, 2011, pp. 705–714.
- [32] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM TKDD*, vol. 5, no. 4, pp. 1–37, Feb. 2010.
- [33] J. Zhang *et al.*, "Social influence locality for modeling retweeting behaviors," in *Proc. IJCAI*, Beijing, China, 2013, pp. 2761–2767.
- [34] P. J. Mucha *et al.*, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, May 2010.
- [35] M. Magnani, B. Micenkova, and L. Rossi. (2013). "Combinatorial analysis of multiple networks." [Online]. Available: <http://arxiv.org/abs/1303.4986>
- [36] P. Shannon *et al.*, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [37] H. Grundmann *et al.*, "Geographic distribution of, staphylococcus aureus, causing invasive infections in Europe: A molecular-epidemiological analysis," *PLoS Med.*, vol. 7, no. 1, p. e1000215, 2010.
- [38] Y. Takhteyev, A. Grudz, and B. Wellman, "Geography of Twitter networks," *Social Netw.*, vol. 34, no. 1, pp. 73–81, Jan. 2012.
- [39] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. ACM SIGCOMM*, San Diego, CA, USA, Oct. 2007, pp. 29–42.
- [40] D. Vogiatzis, "Influential users in social networks," in *Semantic Hyper/Multimedia Adaptation*. Berlin, Germany: Springer, 2013, pp. 271–295.
- [41] R. I. M. Dunbar, "Neocortex size as a constraint on group size in primates," *J. Human Evol.*, vol. 22, no. 6, pp. 469–493, 1992.
- [42] R. Dunbar, "Social networks: Human social networks," *New Scientist*, vol. 214, no. 2859, pp. 4–5, 2012.
- [43] The Economist. (Jan. 2011). *Too Many Friends?*. [Online]. Available: [http://www.economist.com/blogs/theworldin2011/2011/01/sudan\\_facebook\\_and\\_dunbars\\_number](http://www.economist.com/blogs/theworldin2011/2011/01/sudan_facebook_and_dunbars_number)
- [44] MIT Technology Review. (Jul. 2012). *Three Questions for Robin Dunbar*. [Online]. Available: <http://www.technologyreview.com/news/428478/three-questions-for-robin-dunbar/>
- [45] B. Gonçalves, N. Perra, and A. Vespignani, "Modeling users' activity on twitter networks: Validation of Dunbar's number," *PLoS ONE*, vol. 6, no. 8, p. e22656, 2011.
- [46] G. Dezeache and R. I. M. Dunbar, "Sharing the joke: The size of natural laughter groups," *Evol. Human Behav.*, vol. 33, no. 6, pp. 775–779, 2012.
- [47] B. Wellman, "Is Dunbar's number up?" *Brit. J. Psychol.*, vol. 103, no. 2, pp. 174–176, May 2012.
- [48] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. WWW*, Raleigh, NC, USA, 2010, pp. 591–600.
- [49] F. Ronald, K. Ravi, and D. Sivakumar, "Comparing top  $k$  lists," *SIAM J. Discrete Math.*, vol. 17, no. 1, pp. 134–160, 2006.
- [50] F. Mccown and M. L. Nelson, "Agreeing to disagree: Search engines and their public interfaces," in *Proc. JCDL*, Vancouver, BC, Canada, 2007, pp. 309–318.
- [51] M. J. Salganik, P. S. Dodds, and D. J. Watts, "Experimental study of inequality and unpredictability in an artificial cultural market," *Science*, vol. 311, no. 5762, pp. 854–856, 2006.
- [52] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we RT?" in *Proc. SOMA*, Washington, DC, USA, 2010, pp. 71–79.
- [53] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and Twitter social networks," *Comput. Sci.*, vol. 52, pp. 166–176, Mar. 2010.
- [54] J. N. Cummings, B. Butler, and R. Kraut, "The quality of online social relationships," *Commun. ACM*, vol. 45, no. 7, pp. 103–108, Jul. 2002, doi: [10.1145/514236.514242](https://doi.org/10.1145/514236.514242).
- [55] A.-L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, May 2005.
- [56] E. David and K. Jon, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York, NY, USA: Cambridge Univ. Press, 2010.
- [57] T. Antal, P. L. Krapivsky, and S. Redner, "Social balance on networks: The dynamics of friendship and hatred," *Phys. D Nonlinear Phenom.*, vol. 224, nos. 1–2, pp. 130–136, Mar. 2006.
- [58] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: Mining a social network with negative edges," in *Proc. Int. WWW Conf.*, Madrid, Spain, 2009, pp. 741–750.
- [59] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proc. CHI, ACM*, Atlanta, GA, USA, 2010, pp. 1361–1370.
- [60] C. Chew and G. Eysenbach, "Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak," *PLoS ONE*, vol. 5, no. 11, p. e14118, 2010.



**Rongsheng Dong** received the B.S. degree from the China University of Geosciences, Wuhan, China, in 1989.

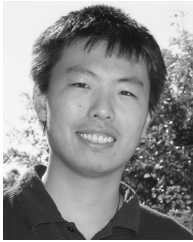
He is currently a Professor with the School of Computer Science and Security, Guilin University of Electronic Technology. He is also currently a member of the Advisory Commission on Teaching Computer Curriculum in Colleges and Universities under the Ministry of Education, the Education Commission of China Computer Federation. He has received research grants from National Natural Science Foundation of China and Natural Science Foundation of Guangxi Province. His current research interests include graph data, social computing, and protocol engineering.



**Libing Li** received the B.S. degree in computer science and technology from the Department of Computer Science, Changzhi University, Changzhi, China, in 2011, and the M.S. degree in computer application technology from the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China, in 2014.

He is currently a Data and Application Architect of Mintoan Insurance Surveyors & Loss Adjusters Group Company Ltd., Shenzhen, China. His current research interests include public sentiment discovery, catastrophic risks modeling, and risk control and loss adjust.





**Qingpeng Zhang** (M'10) received the B.S. degree in automation from the Huazhong University of Science and Technology, Wuhan, China, and the M.S. degree in industrial engineering and the Ph.D. degree in systems and industrial engineering from The University of Arizona, Tucson, AZ, USA.

He was a Post-Doctoral Research Associate with the Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently an Assistant Professor with the Department of Systems Engineering and Engineering Management, City

University of Hong Kong, Hong Kong. His current research interests include social computing, complex networks, healthcare data analytics, semantic social networks, and Web science.

Dr. Zhang is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS.



**Guoyong Cai** received the B.S. degree in applied physics from Hefei Polytechnic University, Hefei, China, the M.S. degree in computer science from the University of Shanghai for Science and Technology, Shanghai, China, and the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China.

He is currently a Professor with the School of Computer Science and Security, Guilin University of Electronic Technology, Guilin, China. His current research interests include social media mining, social computing, and software engineering.