# The Role of Local Influential Users in Spread of Situational Crisis Information

Chao Fan [1], Yucheng Jiang[2], & Ali Mostafavi[1]

[1]Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas, 77843, USA
[2]Department of Computing and Information Science, Cornell University, Ithaca, New York, 14850, USA

*Extensive spread of situational information is important for communities in response to crises/ disasters. Among various mechanisms affecting the spread of information on social media, influential users play a critical role in enhancing information spread. This study examines the attributes and activities of local influential users as well as their interactions with ordinary users on Twitter during 2017's Hurricane Harvey. The results show that the influence across local influential users has a scale-free power law distribution and also indicates a major limitation in spreading information caused by insufficient interaction among influential users themselves. The findings suggest that influential users should play a boundary-spanning and brokerage role in addition to their information hub role in order to be more effective in enhancing the spread of situational information.*

## Lay Summary

*In this study, we used network analysis to examine Twitter traffic during Hurricane Harvey in 2017, focusing on the interaction between local influential users (e.g., individuals with specific knowledge of street inundation, sewage backups, etc.) and more "ordinary" users. The findings are fourfold: first, local influential users who are active early in the cycle of an event tend to gain attention from ordinary users; second, new ordinary users have a preference for users who are already influential; third, a major hindrance to the spread of information to a broader audience was that there were poor links between the influential users themselves; and finally, local influential users build up their influence by being sending out many messages and leveraging this to become "information hubs." We discuss these findings vis-à-vis intervention strategies for officials, emergency responders, and the public with the goal of enhancing the dissemination of situational crisis information via social network sites.*

---

## The role of local influential users in spread of situational crisis information

Disasters lead to community disruptions that rapidly unfold and evolve in a relatively short period of time. Accessing situational disaster information by residents and decision-makers plays a critical role in timely response to disruptions and reduction of impacts on communities (Bagrow, Wang, & Barabási, 2011; Newell, Rakow, Yechiam, & Sambur, 2016). Previous research has shown the importance of information channels in enabling protective action decision-making processes of individuals and households (Lindell, 2018). With the advent of ubiquitous social media platforms such as Twitter and Facebook, people increasingly utilize these media for information search and communication in crises situations (Zhang, Fan, Yao, Hu, & Mostafavi, 2019). The spread of situational information via online social networks (OSNs), where users can generate, communicate, and share information, could enhance the awareness and collective intelligence of communities in coping with disruptions (Kogan, Palen, & Anderson, 2015; Lu & Brelsford, 2015; Yoo, Rand, Eftekhar, & Rabinovich, 2016). Situational information in this study is defined as the information that would contribute to situational awareness about the disaster event. Hence, it is important to understand and improve the underlying mechanisms that influence information diffusion on OSNs in disaster/crisis settings (Kryvasheyeu et al., 2016).

Researchers from physical, social, and computer sciences have studied different mechanisms (e.g., information content attributes and network properties) related to the spread of situational information in OSNs (Starbird & Palen, 2010). For example, researchers have investigated the semantic features of posts on social media to predict the social amplification of risk communications in OSNs (Sutton et al., 2014). In these studies, the content topics (Vosoughi, Roy, & Aral, 2018) and message styles (Sutton et al., 2015a) were shown to correlate with the popularity of users and the efficiency of information spread on social media (Hutto, Yardi, & Gilbert, 2013). Another branch of studies focused on the structural properties of OSNs for information spread (Weng, Menczer, & Ahn, 2013). The focus of these studies were on modeling the path of information flow based on understanding fundamental structures (such as network motifs) in OSNs (Sarkar, Guo, & Shakarian, 2019). The existing research in this area has shown that these fundamental structural patterns have significant impacts on creation of relationship, inhibition of influence spread, and performance of user groups (Sekara, Stopczynski, & Lehmann, 2016). In addition, by further studying network structures, researchers examined empirical influence models to specify the information diffusion process (Aral & Dhillon, 2018; Morone & Makse, 2015). Two broad classes of diffusion models are commonly adopted in the current literature: linear threshold model (Granovetter, 1978; Valente, 1996) and independent cascade model (Goldenberg, Libai, & Muller, 2001; Kempe, Kleinberg, & Tardos, 2003). Both of these models assume each user contributes independently to the spread of information (Yoo et al., 2016). More recently, linear quadratic models have been proposed for analyzing the influence of neighbors on users' adoption and retransmission of information in social networks (Leng, Dong, & Pentland, 2018). However, little is still known about the varying influence of users on information spread in OSNs.

Social influence is an important factor affecting public opinion (Dubois & Gaffney, 2014; Li & Zhang, 2018), adoption of innovation (Goldenberg, Han, Lehmann, & Hong, 2009), and emotion contagion (Aral, Muchnik, & Sundararajan, 2009; Stieglitz & Dang-Xuan, 2013). On social media, some users (called "influencers" or "influential users") leverage upon their popularity to impose social influence for a particular purpose. The important role of influential users has been demonstrated in various domains, including political science (Weeks, Ardèvol-Abreu, & De Zúñiga, 2017), brand marketing (Goodman, Booth, & Matic, 2011; Vinerean, Cetina, Dumitrescu, & Tichindelean, 2013), and

health promotion (Albalawi & Sixsmith, 2017). In the context of disasters and crises, influential users could play a primary role in spread of situational information (de Albuquerque, Herfort, Brenning, & Zipf, 2015; Kogan et al., 2015; Sutton et al., 2015b; Tkachenko, Jarvis, & Procter, 2017). Local influential users who can gather and communicate local situational information (e.g., local street inundation or sewage backup) become a trusted source of information for affected local users (Liu, Fraustino, & Jin, 2016; Peters, Covello, & McCallum, 1997). In fact, some local influential users gain significant popularity because of their activities during a discourse of a disaster and become community champions/celebrities (Yang et al., 2019). In addition to individual's activities on social media, existing studies (Doerfel, Lai, & Chewning, 2010) have also investigated how organizations integrate the Internet into crisis communication. For example, Perry, Taylor, and Doerfel, (2003) showed that many organizations are turning to the Internet to communicate with the public and the news media during a crisis. More recently, studies (Chewning, Lai, & Doerfel, 2012) revealed that organizational use of information and communication technologies helped aid recovery after Hurricane Katrina and gained popularity for organizations. Hence, both individuals and organizational users engage in communications on social media, gaining public attention and helping communities cope with natural disasters.

While popularity and recognition by the community could be indicators of the importance of online users in disasters, the characteristics of these users and the specific role they play in the spread of situational information is not fully known. A key question persists: are influential users merely disaster celebrities or do they play a more significant role such as being community experts (information hubs) or information brokers in times of crises? In this study, we present an analysis related to the spread of situational information by local influential users using Twitter posts collected during the 2017 Hurricane Harvey in Houston. Three research questions investigated in this study are proposed as follows:

1. Who are the local influential users in public warning and situation awareness on social media during disasters?
2. To what extent do the information sharing activities of local influential users vary from those of local ordinary users? To what extent do their behaviors contribute to the spread of situational information in disasters?
3. What role do local influential users play in spreading situational information in OSNs?
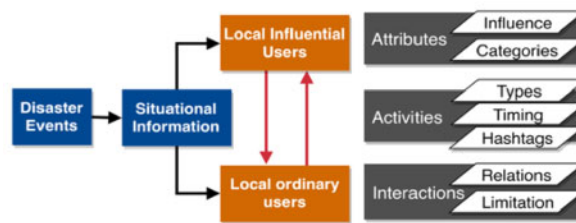
The context of the study is the 2017 Hurricane Harvey in late August in Houston, Texas. Hurricane Harvey was the first major hurricane, attained Category 4 intensity, to make landfall in the United States since Wilma in 2005 (Mooney, 2017). To complement official response and relief efforts, local users increased their activities on social media by posting, sharing, and commenting on situational information related to disruptions, relief needs, and early warnings.

## Methods

In this section, we discuss the details of the methods to perform our analyses (Figure 1).

### Data collection

The data related to Hurricane Harvey were collected through Twitter PowerTrack application programming interface (API) from Gnip, a data provider. The data include the text of the tweets, posted time stamps, retweet/reply status, URLs to external sources, as well as users' profile information such as user profession, locality, profile creation date and follower count. The collection period extends

**Figure 1** An analytical framework for exploring the mechanism of spreading situational information by local influential users.
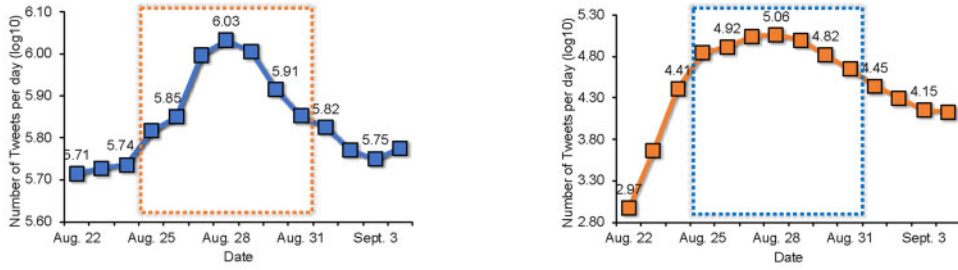
from mid-August 2017 until the end of September 2017, which covers the periods before, during and after Hurricane Harvey. To focus on the communication among local users, we defined geographical boundaries which cover the area of Houston to filter geotagged tweets and created a filtering rule that enabled us to grab all the tweets posed by users whose profile localities were in Houston. With the temporal and geographical rules, we collected 21 million tweets from 700,000 local users over Houston area during that time.

Data used for this study were filtered according to their content and relevance to the situation during Hurricane Harvey, using a preselected set of keywords: "hurricane," "Harvey," "flood," "storm," "disaster," "relief," "i10," "sh6," and "reservoir." The first six keywords are general keywords related to Hurricane Harvey, and the rest of the keywords are important infrastructure such as interstate highway 10 ("i10"), state highway 6 ("sh6"), and Addicks and Barker reservoirs ("reservoir") that were disrupted during Harvey, severely impacting nearby residents. After filtering the tweets with situational content from Aug. 25 to Aug. 31, we obtained 586,010 related tweets posted, shared or quoted by 44,216 local users.

As shown in Figure 2, the volume of tweets generated by local users greatly increased and stayed at a high volume of daily activity from August 25 to August 31, 2017. The peak number of general tweets posted by local users is about 6 million, appearing on August 28, while the peak number of tweets showing situational information is about 0.5 million appearing on the same date. To study the spread of situational information and the role of local influential users, we collected all tweets containing situational information and examined these local users to identify the influential ones to evaluate their activities.

## PageRank for measuring user influence

User communication behaviors including retweeting, replying and quoting a tweet posted by a user are considered as a form of social endorsement (Metaxas et al., 2015). It should be noted that as we observed in the context of disasters, replying mainly focus on requesting or sharing more situational information to improve the awareness of the situation. In this study, we also considered replying as a form of endorsement for spread of situational information. Social influence of online users is the result of endorsement activities led by other users in OSNs. The more the users are endorsed in online communications, the more influential they could be. As documented in relevant literature, the communication activities of users on social media can be converted to an endorsement network (a.k.a. retweeting network) (Stella, Ferrara, & De Domenico, 2018). In this study, replying and quoting is also considered as a retweeting behavior because Twitter displays the replies and quotes as a same form of retweets, which has the same function of sharing information and endorsing statement.

(a) General content by local users

(b) Situational information by local users

**Figure 2** The temporal distribution of the tweets posted by local users. The dots that are included in the rectangle are the selected dates for this study.

The retweeting network in this study includes all local users (44,216 users) and their connections created by retweeting behaviors. The links between the users have directions from the user who retweets the content to the user who posts the content. Because a hurricane is a rapid-onset disaster which attacks the areas intensively but only lasts a few days, the influence of the users during this short time period (comparing to studies in long-term communication) is considered to be stable. Hence, in this study, we created the OSN based on all the tweets posted during the hurricane period, and the analyses are conducted based on this network. Then, we adopt PageRank, which has been widely used as an approach for identifying super spreaders, as a measure of user influence in the retweeting network. There are numerous metrics to measure social influence in OSNs, such as PageRank, degree and betweenness (Gibson, Kleinberg, & Raghavan, 1998; Teng, Pei, Morone, & Makse, 2016). All these metrics have been demonstrated as good predictors of nodes' importance in a network. PageRank is a method for quantifying how important individual nodes are for information flow in a given network topology, and is widely used in webpage ranking and user influence (Stella et al., 2018). Consistent with the existing literature (Riquelme & González-Cantergiani, 2016), this study adopted the PageRank method to measure the influence of users in Twitter network. The PageRank algorithm is based on the idea of random walk which starts with a random node and moves to other nodes according to the probability of a link between them. The transition matrix is shown as follows:

$$M_{ij}' = \beta M_{ij} + (1 - \beta)\frac{1}{n} \tag{1}$$

where $M_{ij}$ is a column stochastic matrix in which columns sum to 1; $n$ is the number of nodes (a.k.a. users) in a retweeting network. At each step, random suffer has two options: with probability $\beta$, follows a link at random; or with probability $1 - \beta$, jumps to a random node. Common values for $\beta$ are in the range 0.8 to 0.9. Once we define the matrix $M$, the PageRank of a user in the retweeting networks can be obtained as follows:

$$r = M' \cdot r \tag{2}$$

where $r$ is the PageRank of a node. The algorithm runs until it converges (i.e., $\sum_j r_j^{(t)} - r_j^{(t-1)} < \varepsilon$, where $t$ is index of iteration and $\varepsilon$ is a predefined small error). This approach helps us get the quantitative influence scores for each local user who was active during the period of Hurricane Harvey.

## Annotation of user categories

We selected the top 200 local influential users whose PageRank scores are significantly higher than other local users. The second top 200 influential users have the closest PageRank scores to the first top 200 influential users. Hence, to demonstrate the difference between these users regarding their PageRank scores, we first conducted a two-sample t-test to examine the statistical significance between the mean of the PageRank scores of the first top 200 users and the mean of the PageRank scores of the second top 200 influential users. The results are shown in Figure S4 in Appendices. The *p-value* is $1.63 \times 10^{-13}$ which indicates that the PageRank scores of the first top 200 local influential users are significantly greater than the second top 200 users. Then, two researchers independently went through all user profiles in the data set to identify different types of the accounts. Due to the privacy issues, however, Twitter can only allow us to collect limited attribute information from user profiles, including user name, a brief description which is used to identify the category of the users, locality which is used to identify the local users, number of followers and number of followees. Due to the limitations related to collecting other user attributes, we can only examine the gross categories of the users. Based on the statistics of the account types, finally, we defined five categories: individuals, news media, government, business, and others. A large proportion of the users are individuals, and most of the individuals are news reporters and meteorologists. To further specify the composites of each profession, we divided the individual category into three sub-categories: news reporters, meteorologists, and other individuals. Although news reporters and the meteorologists are members of news organizations, their behaviors on Twitter do not represent the behaviors of their official organizations. These users reported their own ideas and situational information gathered from non-organization channels on their personal Twitter accounts. Hence, the authors consider the news reporters and meteorologists as individuals, which are different from the official organizations. The news media category includes the accounts like "ABC13 Houston," "FOX26 Houston," and "KHOU 11 Houston." Government accounts are the accounts that are officially managed by local government, such as "Alert Houston," "Texans PR," and "Houston Police." Business accounts are defined as the accounts operated by commercial agents and industrial companies, such as "Houston Bush Airport," "CultureMap Houston," and "DoubleHorn Photo." Some accounts operated by schools, universities, and non-government/non-profit organizations, or that cannot be specifically identified are categorized into an "Other" category. Two researchers individually annotated the categories for each user, and then compared and discussed their annotation results until they reached 95% agreement on the categories for all users.

## Hashtag co-occurrence networks

Hashtags on Twitter are not only the representations of the main semantic content of a tweet, but also a way of enabling a precise search for relevant tweets. Users include multiple hashtags in their tweets in order to improve the exposure of the tweets to other users in search results. Thus, the association of different hashtags in a tweet reflects a strategy that a user intends to convey and highlight the concepts. In this study, we build the hashtag co-occurrence networks in which the nodes represent different hashtags and the links represent the co-occurrence of two hashtags in a tweet. The links are also weighted by the frequencies of the co-occurrences of two hashtags. We build two hashtag networks upon our Twitter data: one for the posts generated by local influential users; and one for the posts generated by ordinary local users.

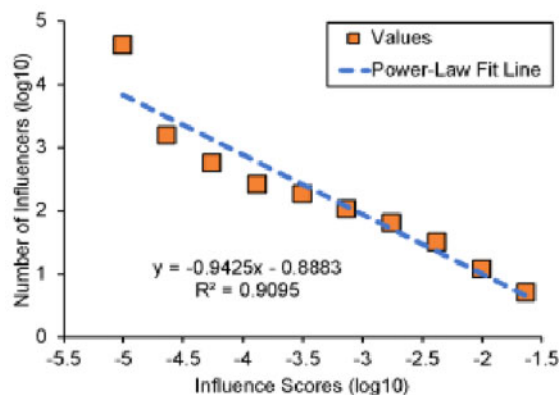## Simulation of information diffusion in OSNs

Various models such as linear threshold model and independent cascade model exist to simulate the information diffusion in social networks. Here, we leverage a simple greedy algorithm to go through the neighbors of top 1000 local influential users and examine the cumulative number of reachable nodes in the retweeting network. The reason for choosing the top 1000 influential users are twofold: first, the influence of these users are significantly greater than other users (the result of two-sample t-test is shown in Figure S5, and the *p-value* is $1.47 \times 10^{-21}$, which indicates the influence of top 1,000 users are significantly greater than other ordinary users); and second, it is more efficient to conduct the greedy algorithm since the algorithm is time-consuming. We also did analysis by choosing different number of influential users such as 500, 1,500 and 2,000. The results for these tests are shown in Appendices Figures S1 and S2. The idea of the simple greedy algorithm is that we activate the most influential users at time $t = 0$. Then, we activate the neighbors of this most influential user along the reversed direction of the links at time $t = 1$. Next, we activate the neighbors of the activated users at time $t = 2$. We repeat this process until all nodes with in-degree links in the network are activated. The algorithm starts with the most influential users and goes to the neighbors of the influential user along reversed directions of the links until it reaches to the nodes without any extra out-degrees. All nodes reached by the algorithm in the OSN through the influential user are considered as reachable nodes for a particular influential user. Then, we seed the information to the second most influential user and rerun the algorithm to identify its reachable nodes. By removing the overlap between two sets of reachable nodes for each local influential user, we can calculate the number of reachable users when seeding information to one more local influential user to the OSN. This measure is defined as cumulative number of reachable users. We iterate the algorithm until all the top 1000 local influential users are added.

## Results

### Attributes of local influential users

The main characteristics of local influential users are examined from two aspects: attributes including number of followers and account categories; and activities including posting, sharing, and replying activities. In this section, we focus on the attributes of local influential users to uncover their general attribute patterns. We quantitatively measure the influence of a user using PageRank algorithms. Figure 3 shows the distribution of local users across the influence scores. Among all users in this study, around 93% of the users were ordinary users whose influence score was around $10^{-5}$, meaning they have a small influence on their neighbors in OSNs. Less than 10 users had influence scores of $10^{-1.5}$, which are considered as highly influential users.

Using log-log plot (Figure 3), we found that the distribution of the influence scores follows a weak "power-law" distribution (using least squares based optimization because Maximum-Likelihood Estimation for power law distribution is not strictly defined for exponents smaller than 1 (Noulas, Scellato, Lambiotte, Pontil, & Mascolo, 2012)). This finding indicates that the structure of OSNs related to communicating situational disaster information is scale free. Hence, this finding implies the existence of preferential attachment in information spread process in OSNs during disasters. That is, new users joining the communication of situational information tend to connect to the influential users. Accordingly, preferential attachment leads to the increasing popularity of local influential users. This finding is consistent with recent studies suggesting that the number of

**Figure 3** The distribution of local influential users over the influence scores.
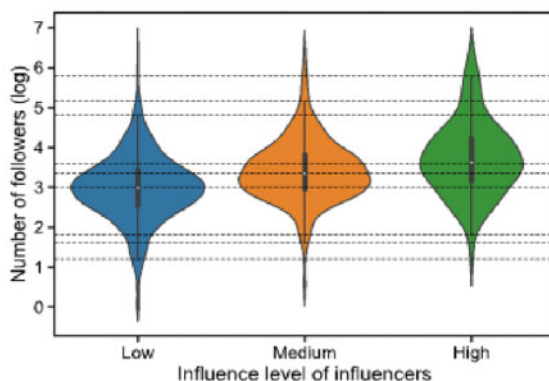
followers is an indicator for identifying influential users in regular daily situations (Cha, Haddai, Benevenuto, & Gummadi, 2010). Aside from the 93% local ordinary users, we filter the remaining 7% local users (2695 local users in total with relatively higher influence) to study the relationships between the level of influence and the number of followers in the context of disasters. As shown in Figure 4, in general, the greater the influence score, the more the number of followers. However, this is not always true. The number of followers for most influential users, regardless of their influence, ranges from $10^3$ to $10^4$. Users with fewer followers could also have high influence in crises/disasters. Hence, the number of followers as an attribute cannot be the only determinant of user influence, although it indicates a general positive relation with user influence.

Another attribute is the categories of local influential users. In a regular situation, the categories of influential users may vary from news media, to celebrities, to politicians. In disasters, however, the categories of influential users might be different because local users have a collective focus on local damages and relief. Understanding the categories of local influential users in a disaster setting can inform the typology of users and its effect on their influence. We annotated the categories of the top 200 local influential users and identified the following: news media, government, business, individuals, and other organizations (Figure 5). In this case, most of the local influential users are individuals (and not official accounts of public or private organizations). This finding indicates that the identity of individuals and their expertise could be an influencing factor in gaining attention needed for ordinary users who seek for information. Among the individual influential users, most of them are news media reporters and meteorologists. In fact, the portion of individual news reporters is much greater than the official accounts of news media in the list of top local influential users. This confirms the importance of expertise in gaining attention needed to become an influential user.
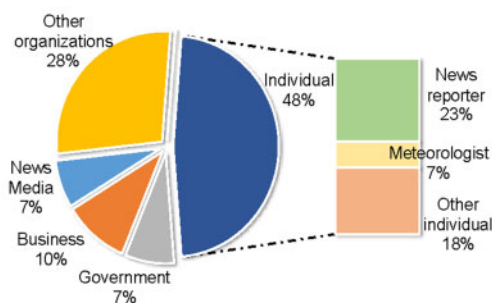
### Activities of local influential users

In addition to the attributes of local influential users, we hypothesize that activities (such as the ratio and timing of posting, sharing, and replying activities) of local influential users also contribute to their influence and the performance of information spread in OSNs. To this end, we examined the distribution of the users with different ratio of posting, sharing and replying activities for both local influential users and ordinary users (Figure 6). The Figure 6 is a ternary plot, which shows the density
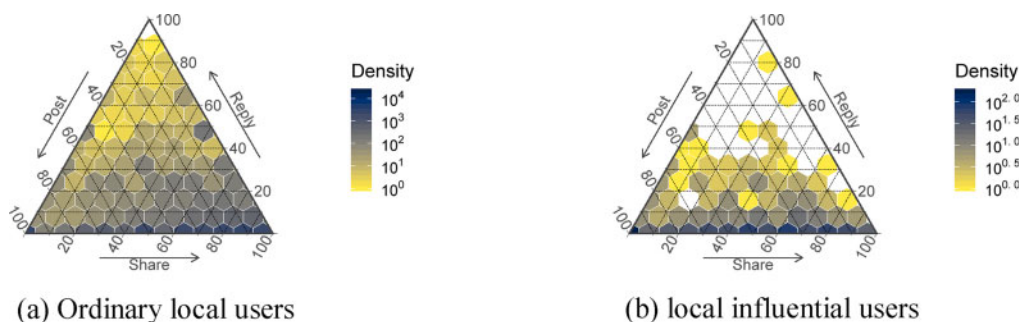
**Figure 4** The relationship between the number of followers and influence levels. We define three influence levels: low; medium; and high. Low influence is corresponding to the influence scores in the range from $10^{-4.8}$ to $10^{-4}$; medium influence corresponds to the influence scores in the range from $10^{-4}$ to $10^{-3}$; and high influence corresponds to the influence scores in the range from $10^{-3}$ to $+\infty$.



**Figure 5** User categories of the top 200 local influential users (see Methods for the definition and examples of the user categories).



(a) Ordinary local users                                    (b) local influential users
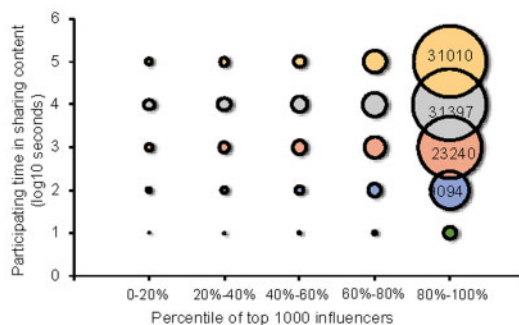
**Figure 6** The differences of activities between ordinary local users (users except the top 1,000 local influential users) and local influential users (top 1,000). "Post" represents the original tweets or quoted tweets posted local users; "Share" represents the retweets which retweet someone's posts; and "Reply" represents a tweet that is created to reply to a post.

distribution of users with different proportions of user behaviors (e.g., posting, replying and sharing) on Twitter. Each hexagon represents the number of users with a certain mixing pattern for the three behaviors. The proportions of the behaviors for the users in each hexagon sum to 1. The values of the gridlines are corresponding to the numbers with the same directions on a side of the triangles. This figure shows that there are a number of ordinary users who had a large proportion of replying behavior (the top corner of the triangle is densely distributed by users) and a large proportion of sharing behavior (the bottom right corner of the triangle is densely distributed by users). However, the majority of local influential users focused on posting and sharing behaviors (the bottom left corner and bottom of the triangle are densely distributed by users). That is, local ordinary users spread the situational information primarily through sharing activities (a.k.a. retweets). Some of them also focused on replying to the posts generated by other users. Unlike the ordinary users, the prominent activity pattern of local influential users is that they tend to post or share information and rarely reply to others' posts. This finding indicates that local influential users strengthen their social influence by generating and sharing information relevant to the needs of local ordinary users. Local influential users do not communicate or respond to requests for relief. Hence, they act more as information hubs (Gibson et al., 1998; Weber & Monge, 2011).
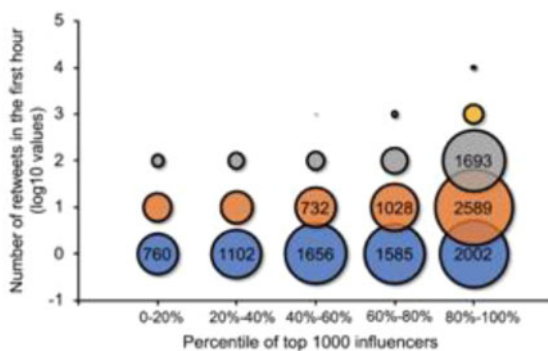
Having found that posting and sharing activities are the primary activities of local influential users, we explored how efficiently influential users share situational information and how effectively their posts are propagated by local ordinary users. We adopted the logarithmic lag interval to measure how long influential users retweet the situational information after they were originally posted by other users. This measure has been applied in existing studies (Shao et al., 2018) and contribute to answering when influential users tend to get involved in the spread of situational information. Figure 7 shows logarithmic lag intervals for the top 1000 local influential users in sharing information, and the relations between the logarithmic lag interval and their influence scores. The result suggests that the top 200 local influential users are more active in sharing posts than low-influence and ordinary users (*p-values* in the table are lower than 0.05, see Figure S1 in Appendices). Among the top 1,000 influential users, most of the retweets were created by top 200 local influential users. In general, it takes less than three hours ($10^4$ seconds) for local influential users to share a tweet posted by other users. The higher the influence of a local user, the less the lag interval of an influential user getting involved in sharing situational information. This finding implies that top influential users play a key information brokerage role in addition to information generation. The combination of information generation and brokerage increases the influence of top influential users.

As documented in existing literature (Shao et al., 2018), the efficiency of spreading situational information for local influential users is also reflected by the number of retweets in the first hour after a post was generated by a local influential user. As shown in Figure 8, we find that high-influence users (i.e., the top 200 influential users in this analysis) can gain more retweets in the first hour than other local influential users. Some posts generated by top influential users can even gain 10,000 retweets in the first hour. However, not all tweets from influential users spread with that speed and magnitude. Seventy percent of the posts generated by local influential users only get less than 10 retweets in the first hour. It is also the case for local influential users with different levels of influence threshold (see Supplementary Materials Figure S1). The lower the influence of a user, the lower the speed of his/her posts being retweeted.

Another important factor related to the activities of the influential users is co-occurrence of hashtags. Hashtags are important features of tweets that help ordinary users seek and find relevant information. Co-occurrences of hashtags in user activities not only contribute to enhancing the exposure of information on Twitter search platforms, but also can highlight the semantic content of the tweets
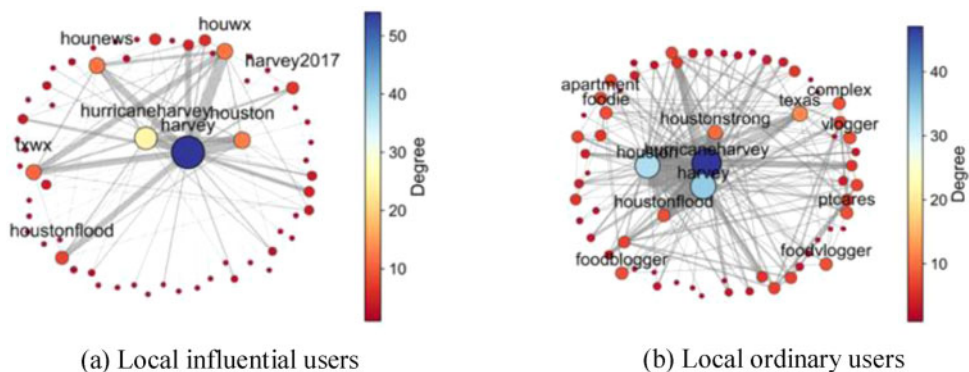
**Figure 7** The logarithmic lag intervals in sharing situational information of top 1,000 local influential users. The higher percentage means the higher social influence of the local influential users. "Grey" bubble represents the average lag interval is about three hours; and "Yellow" bubble represents the average lag interval is about one day.



**Figure 8** The number of retweets in the first hour when a tweet was generated by local influential users. The higher percentage means the higher social influence of the local influential users. The "dark blue" bubble represents the number of retweets created in the first hour is close to 0; and the "orange" bubble represents the number of retweets created in the first hour is close to 10, so on and so forth.

to catch other users' attention and retweets (Stella et al., 2018). To characterize the patterns of using hashtags used by local influential users, we build and analyze the networks of hashtag co-occurrences, providing evidence of users' semantic focuses and strategies for increasing the exposure of their posts (Figure 9). Unsurprisingly, the hashtags "#hurricaneharvey," "#houston," and "#harvey" are the most popular hashtags in both networks. These three hashtags also have strong association relations with each other. In addition to the similarities, Figure 9 shows the extent to which the same hashtags co-occur differently by various users. That is, some semantics are important in the posts made by one type of users but peripheral in the posts by other users. For example, local influential users use less diverse hashtags and use hashtags that have strong relation with the names of the disaster-hit locations and disaster events, such as Houston, Harvey, and Texas; while local ordinary users employ diverse hashtags that are not semantically related to Harvey. Hashtag, "#houstonstrong," was popular among
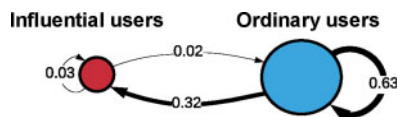
Figure 9 Hashtag co-occurrence networks in the tweets posted by local influential users or ordinary local users: Nodes represent hashtags, edges represent co-occurrence association, and weights of edges represent the frequency of the co-occurrence in original posts (do not consider retweets, replies and quoted tweets).

the local ordinary users and has a strong association with other disaster-related hashtags. This hashtag carries some emotions, which shows encouragement words by users for the disaster-affected people. As suggested by the existing studies (Hwong, Oliver, Van Kranendonk, Sammut, & Seroussi, 2017), hashtags in Twitter tend to represent the topics of the tweets. The results presented in Figure 9 are obtained from quantitative measures (number of co-occurrences and node degree). The nodes labeled in the hashtag co-occurrence networks are the nodes with high degrees. Specifically, the examples in Figure 9 indicate that local ordinary users associated #Harvey frequently with diverse hashtags such as #foodblogger, #houstonstrong and #apartment, while influential users mainly focus on disaster-related hashtags. This result indicates local ordinary users tend to post tweets with their reactions while local influential users consistently focus on posting situational information throughout the affected area.

### Interactions between local influential users and ordinary users

Based on the above findings, it is evident that local influential users have a key role in information spread in OSNs through interactions with local ordinary users. To better characterize the nature of the observed interactions between local influential users and ordinary users in OSNs, we investigate the proportion of their interactive activities (e.g., retweets of influential users posts by ordinary users and vice versa). Figure 10 summarizes the structure of influential-ordinary user interactions (retweeting behaviors). While ordinary users interact mostly with other ordinary users (63% of the retweets), 32% of the retweets are made by ordinary users on the posts of influential users; however, only 2% of the posts by ordinary users get retweeted by influential users. This result indicates that influential users do not rely on ordinary users as an information source (while ordinary users rely on influential users as information hubs). Based on their expertise, influential users have their own information sources for reporting situations in disasters. Meanwhile, interactions among local influential users are not extensive either. This result suggests that the focus of local influential users on information

**Figure 10** The interactions between local influential users and ordinary local users. The sum of the numbers on four links are 1, showing the proportion of each type of interactions among different types of users.
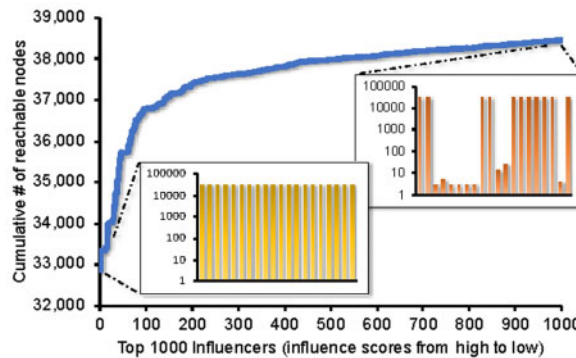
generation in their own domain expertise limits their boundary spanning activities (spreading information posted by other influential users with other expertise).

The limited interactions among local influential users and extensive interactions between influential users and ordinary users is a contributing mechanism to the preferential attachment process for user influence identified earlier. To shed light on the nature of these user interactions and the effects on information spread in OSNs, we examine the networks through a simulation of information seeding and diffusion (See Methods for details). Figure 11 shows that the top 1,000 influencers can reach around 37,500 local users in OSNs, which accounts for 84.8% of the local users in the entire data set. The top 200 influencers can reach 37,000 local users, accounting for 98.7% among the total reachable users by top 1000 influencers. Thus, the top 200 local influential users can reach to almost all reachable users identified in the neighbors of top 1,000 local influential users. In other words, efficient information spread could be achieved by having a certain number of influential users with varying level of influence. After a threshold for the number of influential users, additional influential users do not significantly improve information spread.

The analysis shows that the effect of seeding information to more influential users on information spread (measured by the number of reachable nodes) diminishes with the addition of seeded influential users until a certain threshold is reached. This threshold is 200 in the case of OSN in Houston metropolitan area during Harvey. This threshold could be a function of size and structure of an OSN, as well as the characteristics of local influential users. The analysis of information diffusion process reveals that there are only a few short plateaus with significant overlaps in the reachable nodes for top 200 local influential users. This result also implies that interactions among top local influential users were not extensive. Interestingly, as information hubs, local influential users hardly interact with each other in the existing retweeting network. It is also the case for local influential users with different levels of influence threshold (see Appendices Figures S2 and S3). The limited interactions among local influential users could negatively influence the efficiency of information spread. This could be the cause for the scale-free property of the OSNs and the power-law distribution of influence distribution in OSNs. This finding also suggests the need for more boundary-spanning activities (sharing other influential users' posts) by influential users to enhance the spread of situational information.

## Discussion

This study provides quantitative empirical evidence about the critical role that local influential users play in the spread of situational information in disasters. The study uncovered important characteristics and network features of local influential users that contribute to their influence in OSNs. The

**Figure 11** Cumulative number of reachable nodes by seeding the information to more influential users in OSNs. The subfigures show the number of reachable nodes in the first 20 and last 20 local influential users.

findings in this study could be generalized in understanding the role of influential users in other types of social networks and in regular situations.

One of the key findings is that domain knowledge of users is an important factor for building their social influence in OSNs, and the number of followers is not the only determinant of influence. Specifically, the domain expertise primarily enables local influential users to have more context-related information to post and spread which gains the attention of ordinary users. With the unfolding of multiple community disruptions, ordinary users actively search for relevant situational information. Because of their expertise in gathering and reporting reliable situational information and response strategies, news reporters, meteorologists and government accounts are attractive to ordinary users in disasters. Such temporally growing social influence leads to an improvement in the spreading situational information in OSNs. This finding could be generalizable and be applied to explain the role of influential users in other contexts such as political issues and elections.

Second, the preferential attachment mechanism and scale-free structure in OSNs, observed by a power-law distribution of the influence of local online users, limit the interactions among local influential users. The power-law distribution of the social influence is a consequence of two mechanisms: first, networks expand by the addition of new users (i.e., local ordinary users); and second, the new users attach preferentially to the users who are already influential (Barabási & Albert, 1999). These mechanisms indicate that influential users can attract local ordinary users, but the connections among the influential users are limited. From the perspective of information spread, the limited interaction among influential users could be restrictive. Hence, the information that is posted by a few influential users could be restricted by the limited interactions among them. This problem is reinforced in the context of disasters due to the varying expertise of local influential users. A variety of local influential users spread situational information with respect to different aspects of the disasters such as the weather condition, road closures, and power outages. To deal with this problem, we propose two strategies. One way is to simply encourage the interactions among local influential users, such as retweeting posts from each other and commenting the information posted by other influential users. This strategy enables the creation of direct connections between two local influential users but may suffer some resistances such as influence competitions among influential users. Alternatively, some local ordinary users could play a boundary-spanning role to bridge the disconnection between two local

influential users. For example, an ordinary user can seed the situational information posted by an influential user to another influential user. This would allow ordinary users to create connections with influential users with varying domain expertise.

Third, local influential users build up their social influence by playing a role of information hubs based on their primary activities: posting and sharing, with a focus on strictly relevant situational information. The influence of local influential users affects three information diffusion elements: the speed of sharing the information posted by other users, the speed of getting retweeted by other users, and the usage of hashtags in their posts. Local influential users with greater influence retweet the information from others more efficiently, get retweeted by ordinary users faster, and use hashtags more focused. The information hub effect of influential users contributes to the rapid spread of information and is beneficial for ordinary users in their active information search regarding disaster situations. However, the interactions among these local influential users are limited, which lead to a negative impact on information spread. Understanding the role of local influential users is useful for ordinary users to report situational information to them for retransmission. When multiple local disruptions happen, a limited number of news reporters and meteorologists are not capable of capturing all the situations across the entire disaster-hit areas. Local ordinary users can collect the information in their surrounding areas such as the neighborhood they live in or the road they use for commute. The retransmission of situational information by local influential users can promote the speed of information spread and enable a new source of situational information from ordinary users to contribute to the collective situation awareness in OSNs.

The results presented in this study are reliable and provide new insights due to the uniqueness of the empirical setting and selection of computational experiments. First, we define strict criterions for filtering local users and situational content based on real experience of the authors. The empirical setting, Hurricane Harvey, studied in this article was a unique event due to its scale and the extent of use of social media for information search and dissemination. Hence, the empirical setting is a representative event for large-scale disasters and crises in the future. Second, the methods used for measuring the influence of local users and examining the patterns of user activities are widely adopted in various existing studies. Hence, the results obtained through these data and methods could be generalized to other contexts.

The study also has limits which may influence the generalization of the results to other contexts. For example, hurricanes are a common natural disaster in coastal areas, which has specificities different from other disasters such as earthquakes. The strong wind and intensive rainfall tend to cause large-scale flooding and damages in the affected areas. Compared to other disasters with similar damages such as earthquakes, hurricanes involve situations that evolve over time. People would seek situational information over the entire period of a hurricane. This specificity of the hurricanes would lead to the emergence of influential users who actively and continuously post and share real-time or forecasted situational information. In addition, analyzing the similarities and differences of network dynamics and user influence in various situations would be of interest, which could be investigated in future studies.

This study unveils the phenomenon of information spread intervened by local influential users, which has broad implications to various domains and fields such as politics, marketing, and scientific publications. The attributes and activity patterns of local influential users enable to develop interventions strategies for both officials and individual community members to enhance the situation awareness of local ordinary users to better cope with disaster-induced disruptions. Future studies can adopt the findings in this article to develop real-time detection algorithm to identify potential influencers for seeding and spreading situational information.

## Code for simulation

```
reachable = [] # set of reachable nodes
cumulation = [] # set of all nodes that are activated
total_visited = set()
for node in inf_set: # for loop in the set of influential users sorted by influence scores
    visited = set(node)
    queue = [node]
    count = 1
    while len(queue) != 0:

        temp_node = queue.pop(0)
        adj_nodes = G.predecessors(temp_node) # adjacent nodes (users) in the network

        for n in adj_nodes:
            if n not in visited:
                queue.append(n)
                count += 1
                visited.add(n)

    reachable.append(count)
    for s in visited:
        total_visited.add(s)

    cumulation.append(len(total_visited))
```

Additional Supporting Information may be found in the online version of this article.

## Data availability statement

The data that support the findings of this study are available from Twitter, but restrictions apply to the availability of these data, which were used under license for the current study. Other researchers can submit a request to Twitter to access the data.

## Acknowledgments

## Supporting information

The following supporting information is available for this article:

## References

Albalawi, Y., & Sixsmith, J. (2017). Identifying Twitter influencer profiles for health promotion in Saudi Arabia. *Health Promotion International*, *32*(3), 456–463. https://doi.org/10.1093/heapro/dav103

Aral, S., Dhillon, P. S. (2018). Social influence maximization under empirical influence models. *Nature Human Behaviour*, *2*(6), 375–382. doi:10.1038/s41562-018-0346-z

Aral, S., Muchnik, L., Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21544–21549. doi:10.1073/pnas.0908800106

Bagrow, J. P., Wang, D., Barabási, A. L. (2011). Collective response of human populations to large-scale emergencies. *PLoS One*, *6*(3), e17680. doi:10.1371/journal.pone.0017680

Barabási, A.-L., Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/SCIENCE.286.5439.509

Cha, M., Haddai, H., Benevenuto, F., Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. *International AAAI Conference on Weblogs and Social Media* (pp. 10–17). Atlanta, Georgia, USA: AAAI. https://doi.org/10.1.1.167.192

Chewning, L. V, Lai, C.-H., Doerfel, M. L. (2012). Organizational resilience and using information and communication technologies to rebuild communication structures. *Management Communication Quarterly*, *27*(2), 237–263. https://doi.org/10.1177/0893318912465815

de Albuquerque, J. P., Herfort, B., Brenning, A., Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, *29*(4), 667–689. doi:10.1080/13658816.2014.996567

Doerfel, M. L., Lai, C.-H., Chewning, L. V. (2010). The evolutionary role of interorganizational communication: Modeling social capital in disaster contexts. *Human Communication Research*, *36*(2), 125–162. doi:10.1111/j.1468-2958.2010.01371.x

Dubois, E., Gaffney, D. (2014). The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, *58*(10), 1260–1277. doi:10.1177/0002764214527088

Gibson, D., Kleinberg, J., Raghavan, P. (1998). Inferring web communities from link topology. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space–-Structure in Hypermedia Systems: Links, Objects, Time and Space-Structure in Hypermedia*

*Systems* (pp. 225–234). New York: Association for Computing Machinery. doi:10.1145/276627. 276652

Goldenberg, J., Han, S., Lehmann, D. R., Hong, J. W. (2009). The role of hubs in the adoption process. *Journal of Marketing*, *73*(2), 1–13. doi:10.1509/jmkg.73.2.1

Goldenberg, J., Libai, B., Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, *12*(3), 211–223. doi:10.1023/A:1011122126881

Goodman, M. B., Booth, N., Matic, J. A. (2011). Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal*, *16*(3), 184–191. doi:10.1108/13563281111156853

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*(6), 1420–1443. doi:10.1086/226707

Hutto, C. J., Yardi, S., Gilbert, E. (2013). A longitudinal study of follow predictors on Twitter. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '13*, 821. doi:10.1145/2470654.2470771

Hwong, Y.-L., Oliver, C., Van Kranendonk, M., Sammut, C., Seroussi, Y. (2017). What makes you tick? The psychology of social media engagement in space science communication. *Computers in Human Behavior*, *68*, 480–492. doi:10.1016/j.chb.2016.11.068

Kempe, D., Kleinberg, J., Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining – KDD '03* (p. 137). New York: ACM Press. doi:10.1145/956750.956769

Kogan, M., Palen, L., Anderson, K. M. (2015). Think local, retweet global: retweeting by the geographically – vulnerable during hurricane sandy. In D. Cosley & A. Forte (Eds.), *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 981–993. New York, NY: Association for Computing Machinery (ACM). doi:10.1145/2675133.2675218

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, *2*(3), 1–11. doi:10.1126/sciadv.1500779

Leng, Y., Dong, X., Pentland, A. (2018). Learning Quadratic Games on Networks, 1–15. Retrieved from http://arxiv.org/abs/1811.08790

Li, X., Zhang, G. (2018). Perceived Credibility of Chinese Social Media: Toward an Integrated Approach. *International Journal of Public Opinion Research*, *30*(1), 79–101. doi:10.1093/ijpor/edw035

Lindell, M. K. (2018). Communicating imminent risk. In T. J. Rodriguez H& W. Donner (Eds.), *Handbook of Disaster Research* (pp. 449–477). Cham, Switzerland: Handbooks of Sociology and Social Research, Springer.

Liu, B. F., Fraustino, J. D., Jin, Y. (2016). Social media use during disasters: How information form and source influence intended behavioral responses. *Communication Research*, *43*(5), 626–646. doi:10.1177/0093650214565917

Lu, X., Brelsford, C. (2015). Network structure and community evolution on Twitter: Human behavior change in response to the 2011 Japanese Earthquake and Tsunami. *Scientific Reports*, *4*(1), 6773. doi:10.1038/srep06773

Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., Finn, S. (2015). What do retweets indicate? Results from user survey and meta-review of research. In D. Quercia (Ed.), *Ninth International AAAI Conference on Web and Social Media* (pp. 658–661). Oxford, England: AAAI Press.

Mooney, C. (2017). *The science behind the U.S.'s strange hurricane 'drought' – and its sudden end*. Retrieved from https://www.washingtonpost.com/news/energy-environment/wp/2017/09/07/

Morone, F., Makse, H. A. (2015). Influence maximization in complex networks through optimal percolation. *Nature*, *524*(7563), 65–68. doi:10.1038/nature14604

Newell, B. R., Rakow, T., Yechiam, E., Sambur, M. (2016). Rare disaster information can increase risk-taking. *Nature Climate Change*, *6*, 158–161. doi:10.1038/nclimate2822

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS One*, *7*(5), e37027. doi:10.1371/journal.pone.0037027

Perry, D. C., Taylor, M., Doerfel, M. L. (2003). Internet-based communication in crisis management. *Management Communication Quarterly*, *17*(2), 206–232. doi:10.1177/0893318903256227

Peters, R. G., Covello, V. T., McCallum, D. B. (1997). The determinants of trust and credibility in environmental risk communication: An empirical study. *Risk Analysis*, *17*(1), 43–54. doi:10.1111/j.1539-6924.1997.tb00842.x

Riquelme, F., González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management*, *52*(5), 949–975. doi:10.1016/j.ipm.2016.04.003

Sarkar, S., Guo, R., Shakarian, P. (2019). *Using network motifs to characterize temporal network evolution leading to diffusion inhibition*. Retrieved from http://arxiv.org/abs/1903.00862

Sekara, V., Stopczynski, A., Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(36), 9977–9982. doi:10.1073/pnas.1602803113

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 4787. doi:10.1038/s41467-018-06930-7

Starbird, K., Palen, L. (2010). Pass it on? Retweeting in mass emergency. In *Proceedings of the 7th International ISCRAM Conference* (pp. 1–10). Seattle. doi:10.1111/j.1556-4029.2009.01231.x

Stella, M., Ferrara, E., De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(49), 12435–12440. doi:10.1073/pnas.1803470115

Stieglitz, S., Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, *29*(4), 217–248. doi:10.2753/MIS0742-1222290408

Sutton, J., Gibson, C. Ben, Phillips, N. E., Spiro, E. S., League, C., Johnson, B., . . . Butts, C. T. (2015a). A cross-hazard analysis of terse message retransmission on Twitter. In R. E. Kasperson (Ed.), *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14793–14798. doi:10.1073/pnas.1508916112

Sutton, J., Gibson, C. B., Spiro, E. S., League, C., Fitzhugh, S. M., Butts, C. T. (2015b). What it takes to get passed on: Message content, style, and structure as predictors of retransmission in the Boston Marathon Bombing Response. *PLoS One*, *10*(8), 1–20. doi:10.1371/journal.pone.0134452

Sutton, J., Spiro, E. S., Fitzhugh, S., Johnson, B., Gibson, B., Butts, C. T. (2014). Terse message amplification in the Boston bombing response. In *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management* (pp. 612–621). University Park, PA: The Pennsylvania State University.

Teng, X., Pei, S., Morone, F., Makse, H. A. (2016). Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Scientific Reports*, *6*(October), 1–11. doi:10.1038/srep36043

Tkachenko, N., Jarvis, S., Procter, R. (2017). Predicting floods with Flickr tags. *PLoS One*, *12*(2), e0172870. doi:10.1371/journal.pone.0172870

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*(1), 69–89. doi:10.1016/0378-8733(95)00256-1

Vinerean, S., Cetina, I., Dumitrescu, L., Tichindelean, M. (2013). The effects of social media marketing on online consumer behavior. *International Journal of Business and Management*, *8*(14), 66–79. doi:10.5539/ijbm.v8n14p66

Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. *Science*, *359*(March), 1146–1151. 10.1126/science.aap9559

Weber, M. S., Monge, P. (2011). The flow of digital news in a network of sources, authorities, and hubs. *Journal of Communication*, *61*(6), 1062–1081.doi:10.1111/j.1460-2466.2011.01596.x

Weeks, B. E., Ardèvol-Abreu, A., De Zúñiga, H. G. (2017). Online influence? Social media use, opinion leadership, and political persuasion. *International Journal of Public Opinion Research*, *29*(2), 214–239. doi:10.1093/ijpor/edv050

Weng, L., Menczer, F., Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports*, *3*, 2522. doi:10.1038/srep02522

Yang, Y., Zhang, C., Fan, C., Yao, W., Huang, R., Mostafavi, A. (2019). Exploring the emergence of influential users on social media during natural disasters. *International Journal of Disaster Risk Reduction*, *38*, 101204. doi:10.1016/J.IJDRR.2019.101204

Yoo, E., Rand, W., Eftekhar, M., Rabinovich, E. (2016). Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. *Journal of Operations Management*, *45*, 123–133. doi:10.1016/j.jom.2016.05.007

Zhang, C., Fan, C., Yao, W., Hu, X., Mostafavi, A. (2019). Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management*, *49*, 190–207. doi:10.1016/j.ijinfomgt.2019.04.004