Understanding the Efficiency of Social Media based Crisis Communication during Hurricane Sandy

Abstract

Rapid communication during extreme events is one of the critical aspects of successful disaster management strategies. Due to their ubiquitous nature, social media platforms are expected to offer a unique opportunity for crisis communication. In this study, about 52.5 million tweets related to hurricane Sandy posted by 13.75 million users are analyzed to assess the effectiveness of social media communication during disasters and identify the contributing factors leading to effective crisis communication strategies. Efficiency of a social media user is defined as the ratio of attention gained over the number of tweets posted. A model is developed to identify more efficient users based on several relevant features. Results indicate that during a disaster event, only few social media users become highly efficient in gaining attention. In addition, efficiency does not depend on the frequency of tweeting activity only; instead it depends on the number of followers and friends, user category, bot score (controlled by a human or a machine), and activity patterns (predictability of activity frequency). Since the proposed efficiency metric is easy to evaluate, it can potentially detect effective social media users in real time to communicate information and awareness to vulnerable communities during a disaster.

Keywords: crisis communication; hurricane warning; evacuation; social media; Twitter; hurricane Sandy; disaster management.

1. Introduction and motivation

Extreme weather events have become common in recent decades (NELSON, 2013). Globally, natural disasters cause $520 billion equivalent loss and are responsible for taking 26 million people below poverty line (The World Bank, 2016). Since 1980, the United States alone has spent more than $1.5 trillion for managing 219 weather and climate related disasters which had overall damages exceeding $1 billion (NOAA National Centers for Environmental Information (NCEI) U.S., 2018). Hurricanes along with other natural disasters in 2017 are expected to cause 135 billion US dollar insured cost (Munich RE, 2018). Effective disaster management plays a critical role in reducing the cost of a disaster with implications in its four phases including mitigation, preparedness, response and recovery operations. Households and communities require appropriate resources to meet different needs in these phases of disaster management (Comfort, Ko, & Zagorecki, 2004). To mitigate loss of lives and infrastructure damage, proper preparedness and organized response strategies are crucial. Information availability about the time and severity of an incident can greatly help disaster preparedness, response and recovery operations. Particularly for responding organizations, effective information sharing and coordination are critical (Bharosa, Lee, & Janssen, 2010; Yates & Paquette, 2011). Access to information enhances the efficiency of response actions and increases coordination throughout the network of responding organizations (Comfort et al., 2004).

Online social media platforms facilitate fast and easy exchange of information through sharing, discussion, and communication producing a huge amount of digital content (C.-M. Huang, Chan, & Hyder, 2010). Social media data has been used to investigate many research topics such as human mobility (Hasan & Ukkusuri, 2014; Hasan, Zhan, & Ukkusuri, 2013), transportation (Chen, Frei, & Mahmassani, 2014; Chen, Mahmassani, & Frei, 2017; Ni, He, & Gao, 2017; Rashidi, Abbasi, Maghrebi, Hasan, & Waller, 2017; Zhang, He, Gao, & Ni, 2018), marketing (Alalwan, Rana, Dwivedi, & Algharabat, 2017; Ashley & Tuten, 2015; Misirlis & Vlachopoulou, 2018), tourism and hospitality (Leung, Law, van Hoof, & Buhalis, 2013), public health (Kass-Hout & Alhinnawi, 2013), disaster management (Q. Huang & Xiao, 2015; Kryvasheyeu et al., 2016; Simon, Goldberg, & Adini, 2015) and so on. Risk communication during a disaster plays a critical role in encouraging people to take protective measures (Paton, 2008). The causal sequence of risk communication includes perceived risk to interactive searching to protective actions (Demuth et al., 2018; Mileti & Fitzpatrick, 1992). The reasons why risk communication sometimes does not work as expected include the failure to accommodate the relationship among the complexity of hazards, peoples' lack of experience of hazards, and the need to rely on others to acquire pertinent information (Paton, 2008).

Online social media can play a vital role in spreading timely updates about emergency and colleting feedback from the affected population. Emergency evacuation plan such as evacuation timing, mode and route choice depends on many different factors (Hasan, Mesa-Arango, & Ukkusuri, 2013; Murray-Tuite & Wolshon, 2013; Sadri, Ukkusuri, & Murray-Tuite, 2013; Sadri, Ukkusuri, Murray-Tuite, & Gladwin, 2014). Information from social media about situational awareness can influence shaping these decisions (Martín, Yago, Zhenlong Li, 2017).

Thus, a wide range of international, state, and local organizations have successfully used social media tools during disasters gaining broader interests among policy makers on how social media might be used to improve disaster response and recovery capabilities (Lindsay, 2011; Reuter & Kaufhold, 2017). For instance, during Hurricane Sandy, social media played an important role by sharing information, when the affected regions had limited access to traditional media (Kaufman, S., C. Qing, N. Levenson, 2012). Most recently, during hurricanes Harvey (Oyeniyi, 2017) and Irma (MacMillan, 2017) many government and local officials have used social media to disseminate information and recovery messages.

When using social media for information dissemination during disasters, it is critical to know what makes an information provider more efficient and effective. However, studying the efficiency and effectiveness of social media users in disseminating information has been a challenging task. Such a study would require appropriate metrics applied over a large collection of disaster communication data. On the other hand, essential components of social communication such as human choices, disaster warning propagation and risk communication in large-scale social networks cannot be reproduced within the limits of typical social experiments.

Increasing use of social media in disasters will require a better understanding of the effectiveness of information spreading to an affected community. Finding the factors of information spreading is crucial for understanding the dynamics of social media systems. A better understanding of the underlying factors will provide insights into effective crisis communication strategies. Users more efficient in spreading information can play an important role during crisis, since user activities can draw a significant amount of attention to relevant topics/content from other users. Understanding the interplay among user activities, network properties and the attention received will help to identify the contributing factors in successful crisis communication in emergency situations (Sadri, Hasan, Ukkusuri, & Cebrian, 2017; Sadri, Hasan, Ukkusuri, & Lopez, 2017). Thus, understanding the influence of social media users has significant implications in a disaster management context. In marketing research, studies (Ashley & Tuten, 2015; Hu & Liu, 2004; Misirlis & Vlachopoulou, 2018) have proposed different metrics for social media based marketing strategies. Although these studies provide insights on how a brand can use social media for customer appeal and engagement, their findings may not be transferable to a disaster context. During a disaster, many agencies, individuals, and influencers participate in information dissemination in a very short period of time depending on the intensity and spread of the disaster. For this reason, efficiency in spreading the information or gaining attention is vital for a successful crisis communication strategy. Although influence of social media users has been studied in many different contexts, efficiency of information/awareness spreading in a disaster context still needs to be investigated.

Previous studies have focused on the popularity of content instead of analyzing the effects of user behaviors on how other users respond to them (Vaca, Aiello, Jaimes, & Milano, 2014). Moreover, to the best of our knowledge, few studies have considered user categories and activity patterns while measuring the efficiency of information spreading in the context of disaster management. In social media dynamics, information diffusion creates sudden bursts of connections (e.g., friends or followers) by creating new edges or deleting existing edges (Myers & Leskovec, 2014). Similarly, during a disaster, information diffusion about situational awareness drives significant changes in the underlying social media connections of friends and followers. Such bursts in new followers may happen due to common interest (textual similarity) (Myers & Leskovec, 2014) or attention (Vaca et al., 2014) to users or information source. In this study, using Twitter data, we analyze the efficiency of social media users in information spreading in the context of hurricane Sandy. We investigate user activity against the attention gained in pre-disaster, during (warning and response phase) and post-disaster periods. The efficiency of a user is defined as the ratio between attention gained over the number of tweets within a period. The research questions and contributions of this work can be summarized as follows:

- Does a more active social media user gain more attention? What combinations of user activity will facilitate such attention in pre-disaster, during disaster and post-disaster periods? This study allows us to understand the correlation between activity and attention gained during these three phases of a disaster.
- How does user efficiency dynamics change over the pre, during and post disaster phases? We show that during disaster specially in hurricane declaration and landfall days users have higher average efficiency than that of pre-disaster and post disaster period.
- What are the factors contributing to user efficiency? How can efficient users be classified based on their activities and features? We present a model to classify efficient users highlighting the features contributing to user efficiency in disaster periods.

2. Background and Related Work

Social media, the computer mediated technology is now one of the most integrated parts of our daily life. These technological advancements have transformed the view of disaster management professionals on disseminating information as well as interacting with affected communities. For instance, previous studies on disaster management found that having a strong social network increases the likelihood of a person responding to a warning message (Aguirre, Wenger, & Vigo, 1998). During crisis, warning message like evacuation decisions can be made anywhere and often with little advanced warning time (Murray-Tuite & Wolshon, 2013). Eye-witnessed information sources provide local and rapid updates during disaster and thus can be more helpful than official news for the decision makers (Palen, Vieweg, Liu, & Hughes, 2009; Shklovski, Palen, & Sutton, 2008).

Previous studies (Kapoor et al., 2018; Shiau, Dwivedi, & Yang, 2017) on social media theories (Ngai, Tao, & Moon, 2015), core knowledge (Shiau, Dwivedi, & Lai, 2018), and applications investigated the behaviors of social media users. In addition, researchers have investigated social media use in disasters from different perspectives (J. Kim & Hastak, 2018; Stieglitz, Mirbabaie, Ross, & Neuberger, 2018). Studying 2013 Oklahoma tornado, it is shown that Twitter data can reveal relevant information as an additional data source for better understanding of individual behavior during a crisis (Ukkusuri, Zhan, Sadri, & Ye, 2014). Visual analytics of microblog data can display public behavior in disaster events (Chae et al., 2014). Mobility patterns can be inferred from geo-tagged tweets (Hasan & Ukkusuri, 2014; Hasan, Zhan, et al., 2013; Sadri, Hasan, & Ukkusuri, 2017). Communities can be detected from user interactions on Twitter (Hasan, Zhan, et al., 2013; Sadri, Hasan, & Ukkusuri, 2017). Social media users can be used as social network sensors to increase disaster awareness (Kryvasheyeu & Chen, 2015). Furthermore, social media data can be used to rapidly assess disaster damage, as it was shown that per capita damages were strongly correlated with per capita twitter activity during hurricane Sandy (Kryvasheyeu et al., 2016).

Activities of social media users are greatly influenced by content production and sharing activities (Vaca et al., 2014). Most of the past studies focused on the popularity or propagation of the content in social media such as popular tweets (Hong, Dan, & Davison, 2011; Mathioudakis, Koudas, & Marbach, 2010), Flickr picture (Cha, Mislove, & Gummadi, 2009), YouTube video (Figueiredo, Benevenuto, Almeida, Fabr, & Almeida, 2011), Twitter hashtag (Lehmann, Gonçalves, Ramasco, & Cattuto, 2012) etc. Scale free networks and affinity affect the propagation of information (Wu, Huberman, Adamic, & Tyler, 2004) but basic measures such as the raw number of social connections are not a good predictor for influence (Asur, Huberman, Szabo, & Wang, 2011; Romero & Huberman, 2011). In addition to the graph properties of user networks, the popularity and influence of a twitter account depend on the personality and emotion of the human being behind that account (Quercia, Ellis, Capra, & Crowcroft, 2011). Stai et al. (Stai, Milaiou, Karyotis, & Papavassiliou, 2018) proposed an epidemic model to understand temporal dynamics of information diffusion in Twitter, explaining the burst like behaviors due to information diffusion (Myers & Leskovec, 2014). Vaca et al. (Vaca et al., 2014) observed that a combination of different type of social and content-producing activity is necessary to attract attention in social media. Using Sina-Weibo data during two typhoons, this study (Dong, Li, Zhang, & Cai, 2018) explores the information diffusion considering individual and network perspective. Analyzing the reposting behavior in Weibo.com during Yiliang earthquake, Li et. al have studied the propagation pattern of different types of information (J. Kim & Hastak, 2018). Kim et. al have analyzed the network characteristics of city of Baton Rouge Facebook page during 2016 Louisiana flood (J. Kim & Hastak, 2018). They have found higher information diffusion in Facebook than Twitter. A study (J. Kim, Bae, & Hastak, 2018) on storm Cindi using twitter data explores the role of four types of twitter users in emergency information diffusion. According to this study, news and weather agencies are the dominant twitter users as information sources whereas the public and organizations are the dominant twitter users as information diffusers. Despite these efforts, what factors contribute to attract attention in social media during a disaster, remains an open question.

In this study, Twitter data before, during and after Hurricane sandy have been analyzed to understand the factors contributing to the overall efficiency of a user in crisis communication. A model is also proposed to classify efficient users based on their attributes. This method has potential to be used to identify effective social media users during disasters for rapid communications.

3. Data

Hurricane Sandy, a late season post-tropical cyclone was the deadliest and most destructive hurricane of the 2012 Atlantic hurricane season**.** On October 20, Sandy's origin was primarily associated with a tropical wave that was assessed as a high potential for it to become a tropical cyclone within 48 hours

(Blake et al., 2013). The hurricane was first classified and officially assigned its name as Sandy on October 22 (Kryvasheyeu & Chen, 2015). After leaving a trail of damage over Jamaica, Cuba, and Bahamas, Sandy made its landfall on the United States at 23:30 UTC on 29 October 2012 near Brigantine, New Jersey. Sandy was responsible for 147 direct fatalities and damage in excess of $50 billion, including 650,000 destroyed or damaged buildings (Blake et al., 2013). Sandy received a lot of media coverage both in traditional media and social media.

The dataset was collected from the publicly accessible data via doi:10.5061/dryad.15fv2 (DRYAD repository) (Kryvasheyeu & Chen, 2015). Kryvasheyeu et al. (Kryvasheyeu & Chen, 2015) collected the dataset through an analytics company Topsy Labs who deals with Twitter data. The collected data contains tweets posted between October 15, 2012 and November 12, 2012. This duration covers the period before the formation of the hurricane to after the landfall in the United States. The dataset contains user id, timestamp, tweet text, tweet id, user followers count, user friends count, sentiment scores and locations. In total, there were 52,493,130 tweets from 13,745,659 unique twitter users. This dataset includes tweets containing some specific keywords relevant to hurricane Sandy and its impact such as ("sandy", "hurricane", "storm", "superstorm", "flooding", "blackout", "gas", "power", "weather", "climate", etc.). The full list of keywords used is provided in Table S1 in the original paper. We use the dataset to understand the efficiency for the hurricane related tweets (all 52.55 million). We have also filtered the raw data with the word "sandy" to get sandy specific tweets. After applying the filter, we obtain around 4.51 million messages from 1.39 million users. We use this filtered data to understand efficiency dynamics for sandy specific tweets.

4. Methods

For the best possible use of social media in crisis communication, it is important to know how an emergency manager should engage with social media users, traditional media, and the targeted population. For crisis communication in blogsphere, Jin et al. (Jin & Liu, 2010) proposed a conceptual model called Blog-Mediated Crisis Communication Model (BMCC). We have adopted and revised this framework as Social-Mediated Crisis Communication Model (SMCC) to incorporate social media (Liu, Jin, Briones, & Kuch, 2012). Influential social media creators/users and social media followers are important components of this social mediated crisis communication model. Our study focuses on an empirical study that are related to the component highlighted by a dotted red box in Figure 1 (left part). This study has three main phases as shown in Figure 1(right part). First, we define efficiency of gaining attention over social media activity. Second, we explore the dynamics of activity, attention and efficiency dynamics in pre-disaster, during disaster and post-disaster period using hurricane Sandy data. Third, we also study the factors that contribute to gain efficiency by adopting Linear regression model and ordered logit model. In the rest of this section we describe the phases of this study in details.

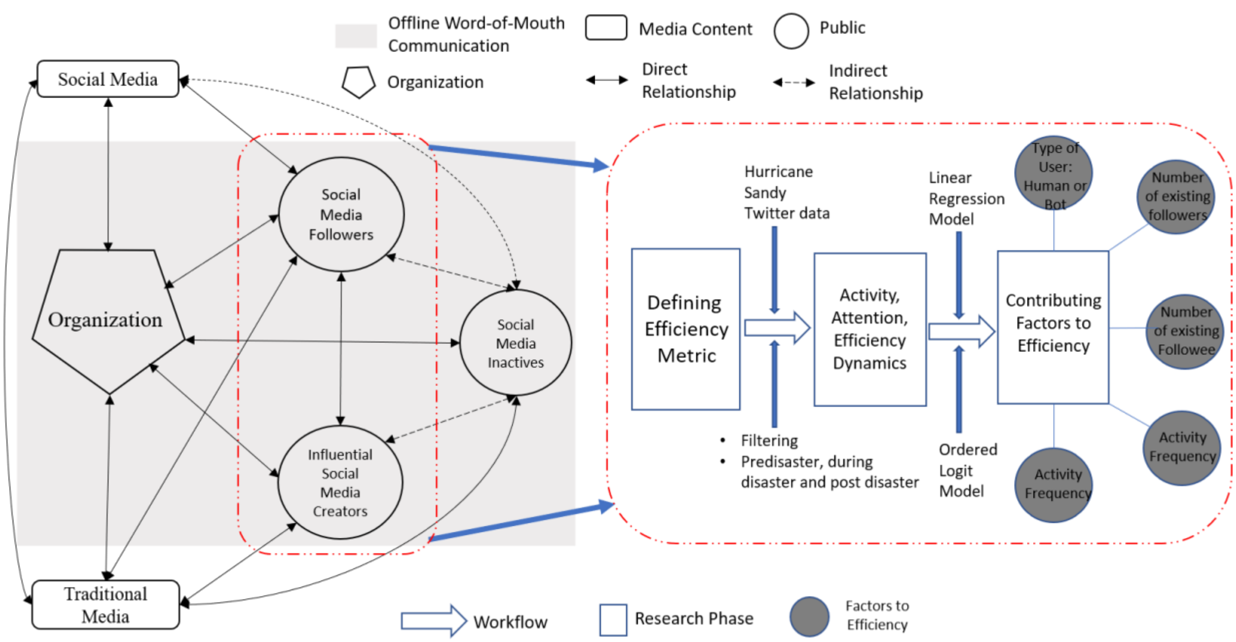


Fig 1. Social Mediated Crisis Communication Model (SMCC) (adopted from (Liu et al., 2012)) **and Study Work Flow**

4.1 Activity, Attention, and Efficiency Metrics in Twitter

We define *attention* as the number of new followers received and *activity* as the number of tweets or new followee added. From the point of view of a particular user, the term "new follower received" means the users that start to follow this particular user, and the term "new followee added" means the users that this particular user starts to follow. For this study, tweet frequency is selected as an activity metric since this is the most frequent activity among all types of users; whereas followee addition is very low or zero for some organizational and personal users. A well-connected user (high initial followers in our study) has a large audience, thus a tweet posted by a well-connected user can reach to many users. But the existing followers may not be the targeted users during an emergency, thus not creating sudden burst in new connections (friends, followers etc.) as study (Myers & Leskovec, 2014) shows that information diffusion creates sudden burst in new connections. For that reason, we have used new follower gain as attention and existing followers as one of the factors. To measure the performance of the users in gaining attention, we use a metric called as *efficiency*. As shown in Equation (1), efficiency η of a user u for the time frame (t_i to t_j) is defined as the ratio between total attention received and total activity performed within that time frame.

$$\eta_u(t_i, t_j) = \frac{\sum_{k=t_i}^{t_j} att_k(\mathrm{u})}{\sum_{k=t_i}^{t_j} act_k(\mathrm{u})} \tag{1}$$

where $att_k(\mathrm{u})$ and $act_k(\mathrm{u})$ represent, in time period k by user u, attention gained and activities posted, respectively.

Although, equation similar to (1) are commonly used in fields like physics and economics, Vaca et al. (Vaca et al., 2014) used this term in a social media setting. Unlike most of the fields where efficiency is upper bounded to 1, it can take any value. Higher efficiency values indicate better engagement and higher influence in social media communication.

4.2 Extraction of User Features

From raw data, for all the tweets of each unique user activity frequency, initial follower count, initial followee count, total follower received, total followee added by the user and efficiency metrics were computed for a selected time interval. To measure the regularity of a user's activity approximate entropy of activities has been estimated. Approximate entropy is a statistical parameter that can quantify the predictability or regularity of a time series data. A repetitive pattern of fluctuation in a time series makes it more predictable than a time series without such patterns. Approximate entropy calculates the likelihood that similar patterns of observation will not found in the data in the subsequent observations. Thus, a higher value of approximate entropy implies less regularity and a smaller value indicates strong regularity (W.-S. Kim, Yoon, Bae, & Soh, 2005). Approximate entropy has been used in many fields such as medical data (Srinivasan, Eswaran, & Sriraam, 2007), finance (S. Pincus & Kalman, 2004), psychology (SM Pincus & Goldberger, 1994), complex system analysis (PINCUS, 1991) etc. Daily activity frequencies of a user for the whole analysis period were used as an input. Approximate entropy was best fitted as it has low computational demand, applicable on small observation (points < 50) and can be applied in real-time. The detailed procedure of computing approximate entropy can be found in this study (Srinivasan et al., 2007). Furthermore, since a significant number of users are being operated autonomously (bot) (Ferrara, Varol, Davis, Menczer, & Flammini, 2014) , we have collected the bot score using truthy botornot-python API to evaluate whether a user account is controlled by human or machine (Davis, Varol, Ferrara, Flammini, & Menczer, 2016).

4.3 Contributing Features to Efficiency

To find the linear relationship between different variables and the outcome variable, univariate and multivariate linear regressions were fitted with the extracted variables. The general form of such models is shown in Equations (2) -(4).

$$Y = \theta_0 + \theta_1 X + \varepsilon \quad (2)$$

$$Y = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3 + \varepsilon \quad (3)$$

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \cdots + \theta_k X_k + \varepsilon \quad (4)$$

Here Y is efficiency, treated as a dependent variable; X, X_1, X_2 etc. are the independent variables affecting efficiency; θ_0 is a constant term; and $\theta_1, \theta_2, \theta_3$ are the coefficients of the corresponding variables. While univariate linear regression describes the relationship of each independent variable with the dependent variable, a multiple linear regression model reveals the relationship of the combined effect of the explanatory variables. The independent variables in the best models are considered as the most influential and explanatory variables in determining efficiency. The best model is selected based on the adjusted *R* squared value.

Users are categorized based on their aggregate efficiency during the whole period. Besides understanding the effect of predictor variables in continuous change in efficiency, we estimate an ordered logit model to understand the effect of the extracted features in predicting the category of the efficiency of a user. An ordered logit or proportional odd model was chosen as the outcome variable is ordered from low efficiency to high efficiency. This model gives the output as the probability or odd of falling an outcome in an efficiency category. The basic equation (Derr, 2013; Torres-reyna, 2012; Washington, S.P., Karlaftis, M.G. and Mannering, 2010) for interpreting this model is given in Equation (5).

$$\log\left[\frac{p_i}{1-p_i}\right] = a_i + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_kx_k \quad (5)$$

$where, p_i = probability\ of\ an\ outcome \leq i$

$a_i = intercept\ for\ outcome \leq i$

b1, b2, etc. are the co-efficient whereas x1, x2, x3 are the explanatory independent variables. The best model is selected based on its AIC value.

5. Results

5.1 Distributions of User Features

This section describes the features collected for a user. Figure 2 shows the distributions of activity, followers followees, bot score, and activity entropy found in the data.

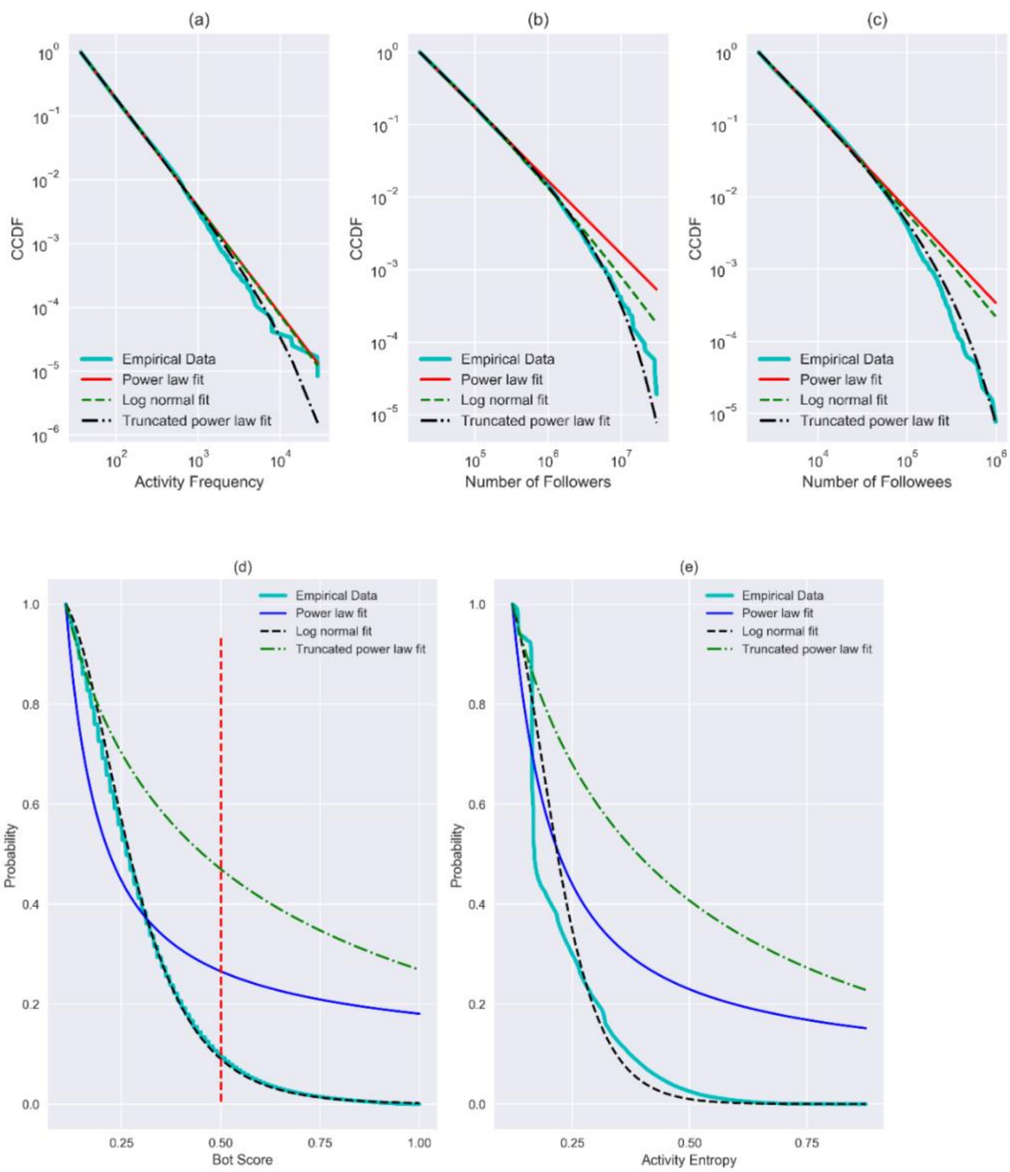


Fig 2. (a) Activity, (b) Follower ,(c) Followee Distributions (d) Bot Score and (e) Activity Entropy. Both X axis and Y axis are plotted in log scale for (a), (b), and (c). Counter cumulative probability (CCDF) is plotted in Y axis which represents the probability of a value x of being greater than the corresponding value in X axis. Both X axis and Y axis are plotted in normal scale for (d) and (e).

Here the follower and followee counts are based on the counts when a user was first observed in the data set. The activity distribution is based on the total number of tweets observed during the whole period in our dataset. Bot score and activity entropy distribution are plotted using values from 646,563 users. A bot score represents the likelihood of being a bot. An extreme value (0 or 1) represents more confidence of the bot-ness of the user. Higher the value, higher the likelihood of being a bot. In our study, we have used 0.5 as a threshold to separate bot-like behavior.

The empirical distributions of activity, initial followers and followees were best fitted to truncated power law among the fitted distributions shown in Figure 2(a), (b) and (c). For bot score and activity entropy, empirical distributions are best fitted to log-normal distribution. Log likelihood ratio tests were used to find the goodness of fit for the fitted power law, lognormal and truncated power law distributions.

5.2 Correlations between Activity and Attention

Activity frequency and followee added play roles in gaining attention. Attention gains may also vary over time and context. Figure 3 shows attention gains for different ranges of activities in pre, during and post disaster periods. Though the duration of pre (9 days), during (10 days) and post (10 days) periods disaster are almost same, followers received is the highest during disaster period compared to other two periods (compare the maximum values of the z scales in Figure 3).

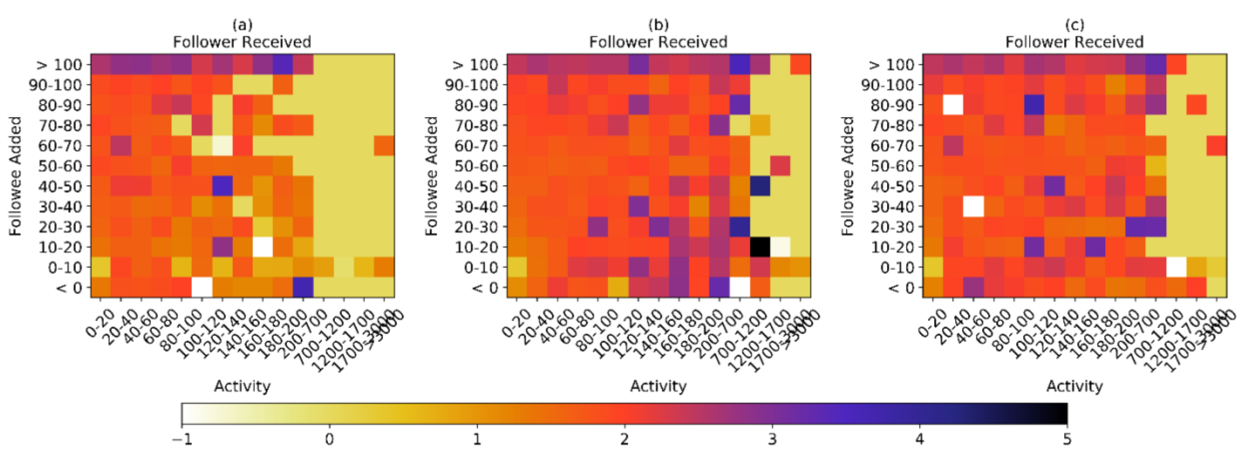


Fig 3. Correlations between Activity and Attention in (a) Pre-Disaster Period (Oct 14, 2012 to Oct 22, 2012), (b) During Disaster (Oct 23, 2012 to Nov 1, 2012) Period and (c) Post Disaster Period (Nov 2, 2012 to Nov 11, 2012). The X-axis and Y-axis represent the range of number of tweets and number of followees, respectively. The Z axis shows the average number of followers received by the users falling in the corresponding x and y bin as color intensity. The Z axis values are shown in log scale.

In general, we do not observe any particular trend between activity frequency and gaining attention in the pre and post-disaster periods. Unlike pre and post disaster periods, during the disaster, higher activities tend to help gaining higher attention. Very high activity frequencies (activity>700) in pre and post disaster periods are not necessarily associated with a high number of followers received. During pre and post disaster periods, users with activity frequency less than 700 have received the highest number of followers (black rectangles in Figure 3); whereas during the disaster, the highest number of followers was gained for a user with activity frequency greater than 700. Followee added less than 100 has no impact in gaining attention but users adding followees greater than 100 have received higher attention during all the three phases. Another observation is that, the highest number of followers received in three phases occurred for the users with activities greater than 100. To study more in depth, we have studied user daily and aggregate efficiency based on different features and categories at different phases.

5.3 User Efficiency Analysis

Daily efficiency of a user is calculated by dividing the total follower gain by the total number of tweets of that day. Similarly, aggregate efficiency is calculated by dividing total daily follower gain by the total tweet in that period. Users are categorized based on their active days. Only the users who had at least one activity on each of the three periods (pre, during, post disaster) are selected in these categories. Figure 4 shows the daily efficiency distribution for the users categorized by their active days. It is found that a significant number of users have daily efficiency value equal to or less than zero. As the number of active day increases, the probability of having a user with efficiency less than or equal to zero decreases. It also shows that the probability of having efficiency less than or equal to zero is maximum for the users who were active less than 8 days. Although, the probability of having daily efficiency greater than 10 is low across all category of users, this probability increases as the number of active days for a user increases.

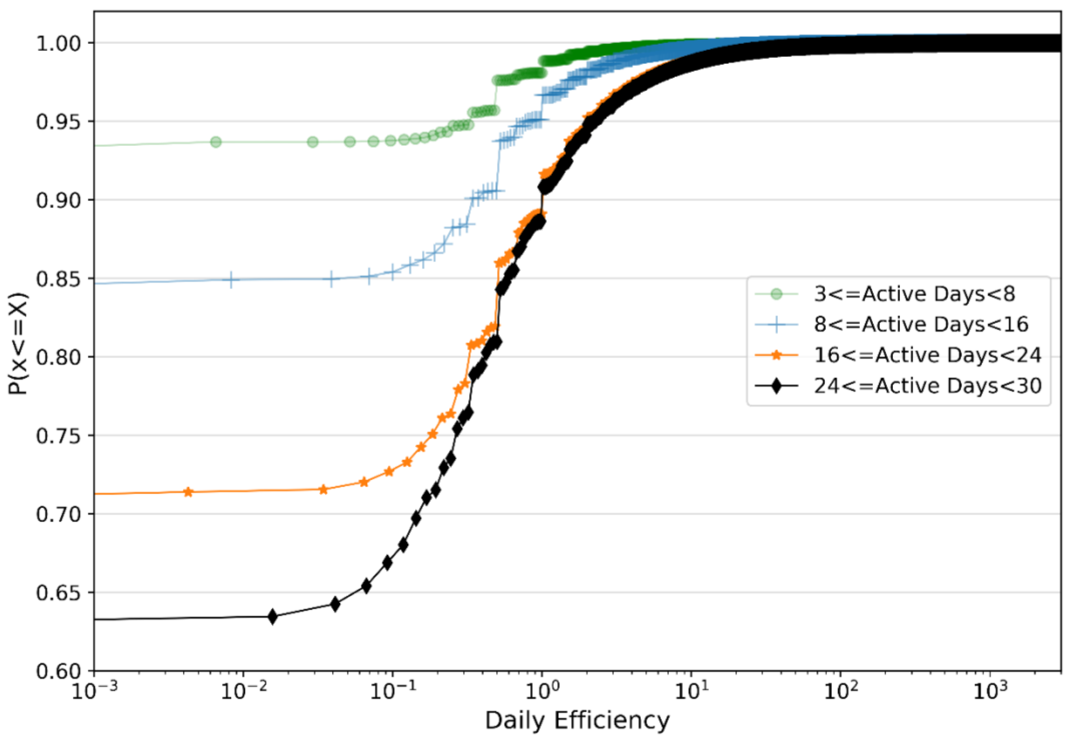


Fig 4. Daily Efficiency Distribution for the Users Categorized by Active Days. X-axis shows the daily efficiency and Y-axis shows the cumulative probability which means the probability of being daily efficiency is equal or less than of the corresponding daily efficiency of X-axis. X-axis are plotted in log scale.

We further investigate, for each user, how daily efficiency varies over time. To find if there is any trend in the data, we have categorized users based on their overall efficiency values (i.e., ratio of sum of daily attention and activity measured over the whole observation period). We calculate average daily efficiency by taking the average of the daily efficiencies of the users for a particular category. For hurricane related tweets, average daily efficiency for the first two categories (overall efficiency less than or equal to zero) does not change that much (see the inset plot in Fig S1). But daily efficiency for the highly efficient users provide an interesting insight. Users had higher values of average daily efficiency during hurricane declaration and landfall days (Fig S1). The spikes on hurricane declaration and landfall days indicate that some users received higher attention for their activities on those days. But this trend shows a significant number of spikes even before the formation of Sandy and also long after its landfall. This indicates that some users might be gaining attention due to tweets unrelated to Sandy. To confirm, we analyze only Sandy related tweets (having 'sandy' within the text of the tweet) and found that efficiency was maximum just after the declaration day (October 23,2012) and decayed readily with a spike at landfall date (see Fig 5). It indicates that users were highly efficient in spreading the awareness about Sandy on the day after its declaration. We do not observe any efficiency curve before declaration because the term 'sandy' was not present before declaration.

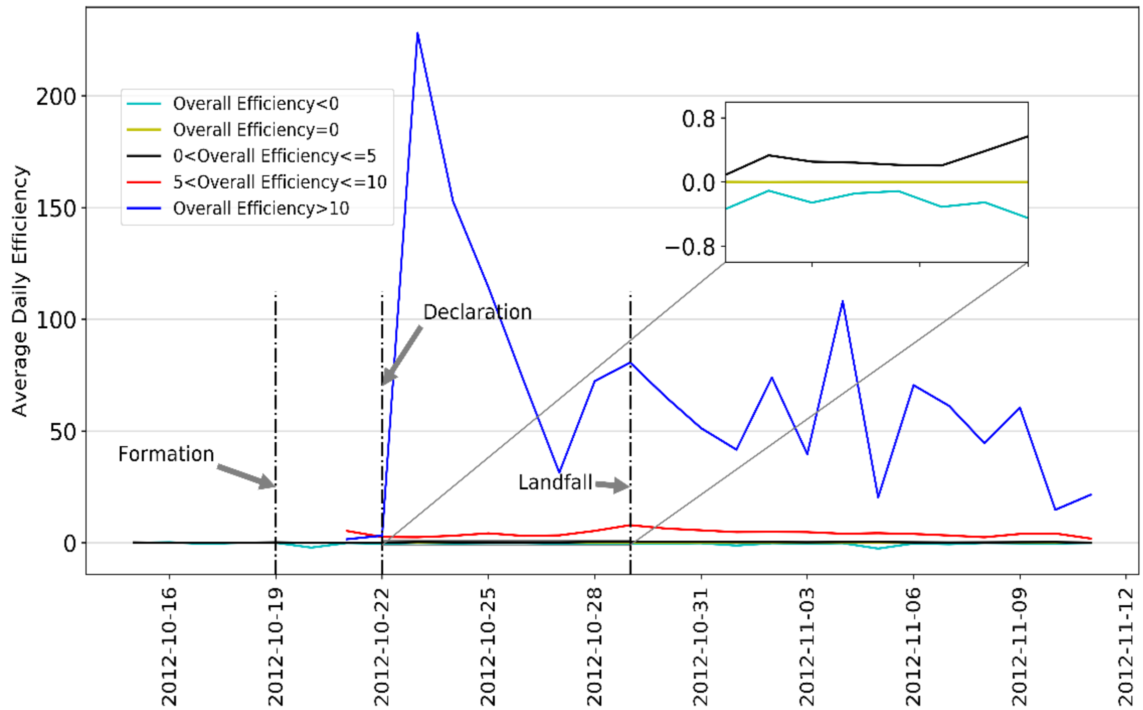


Fig 5. Average Daily Efficiency of the users categorized by efficiency values (for Sandy specific tweets). X-axis shows the time in date and Y-axis shows the average daily efficiency. Users are categorized by their overall efficiency values.

To determine what type of users became highly efficient spreading the awareness during hurricane declaration and landfall, we extract the top 5 efficient users for both hurricane related tweets and Sandy specific tweets. Fig S2 shows the daily efficiency values for the top 5 efficient users on hurricane Sandy declaration and landfall days and top 5 efficient users over the whole data coverage period. We find that majority of them are either political users or have no significant association with hurricane updates (see Table S1). However, analysis on Sandy specific tweets reveals significant spikes close to landfall day (see Fig 6). We find that these highly efficient users are either storm update centers or weather reporter having close association with hurricane updates (see Table S2). It highlights the importance of an appropriate filtering step when identifying highly efficient users, specific to a disaster.

Tweets collected for a general disaster context may contain ambiguous words (e.g., power, weather, recovery etc.) overlapping with other highly conversed contexts.

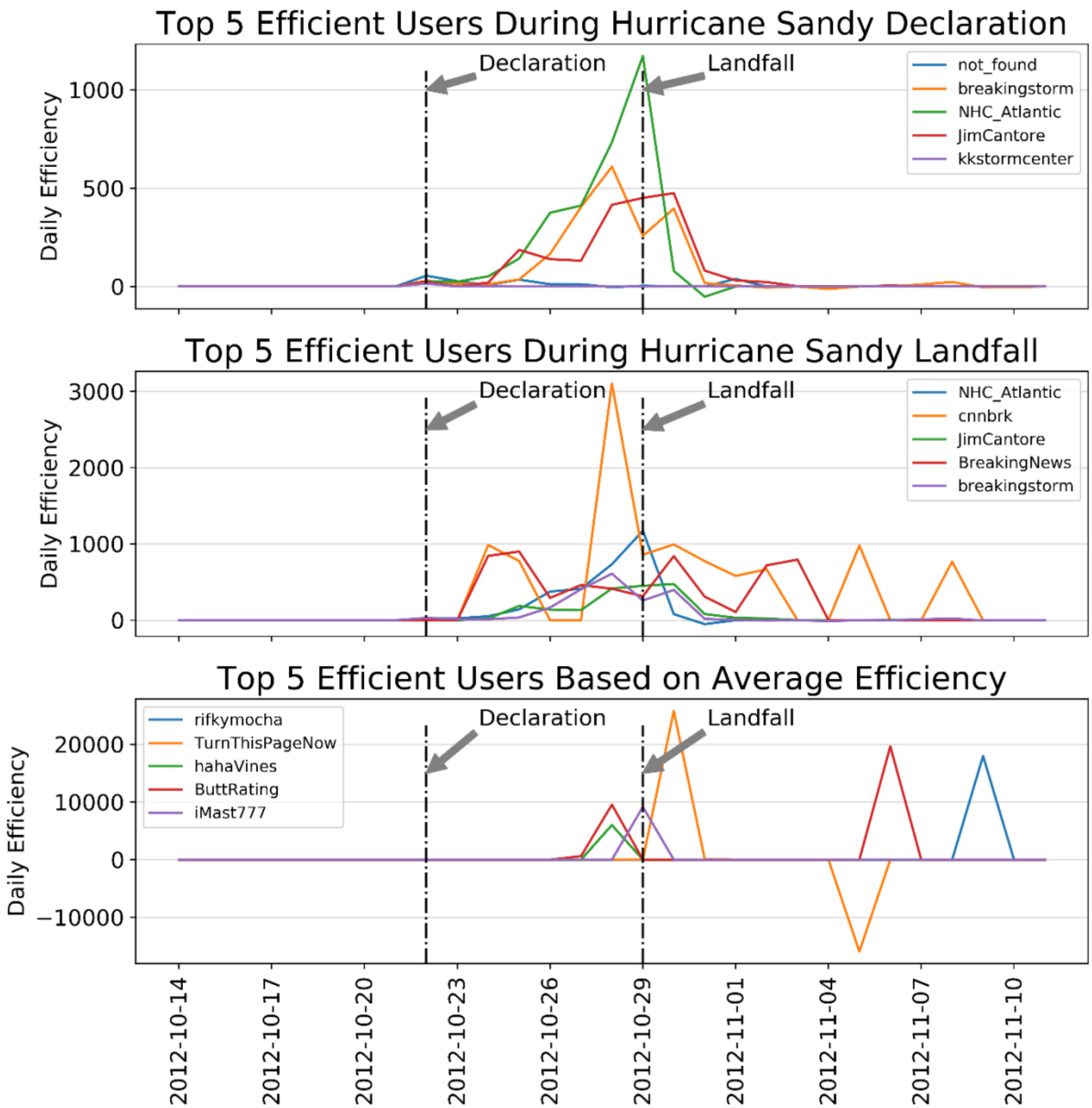


Fig 6. Top 5 Efficient Users during Hurricane Declaration, Landfall and Overall (for Sandy specific tweets)

5.4 User Attributes Contributing to Efficiency

To identify the contributing factors for an efficient spreading of awareness, it is important to know the relationship between each feature and efficiency metric. Similar to the previous section, we analyze separately for hurricane related and Sandy specific tweets for understanding the factors in gaining attention in crisis communication. For hurricane related tweets, Fig 7 shows the relationship between efficiency and each of the variables of followee add, initial follower and initial followee, considering two types of users: bot (bot score >=0.5) and non-bot (bot score <0.5). In addition, the relationship between efficiency and active days has been modeled considering user categories based on activity entropy. The result shows good correlation (R_2 >0.5) between efficiency and initial follower for the non-bot users. From Figure 7, we find that efficiency is positively associated with all the variables. In all cases, R_2 values are lower for bot users compared to non-bot users. This reflects that efficiency of bots cannot be well predicted with a single variable. Similar associations have been found for Sandy specific tweets (see S3 Fig).

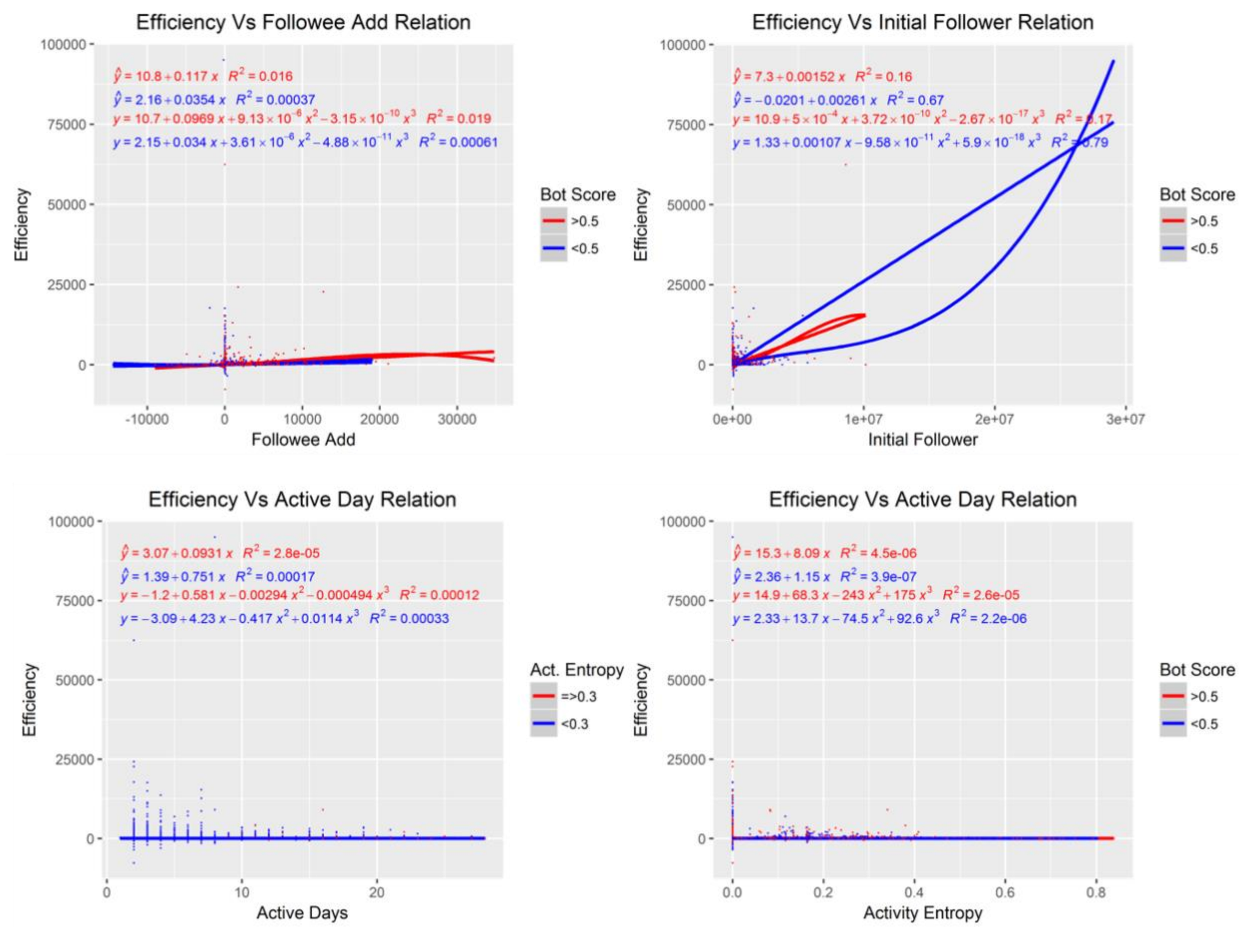


Fig 7. Correlations between Efficiency and Different User Attributes (for hurricane related tweets)

To determine the combined effects of the explanatory variables, we estimate a multivariate linear regression model. We have used an interaction term to capture the additional effect on efficiency of a non-bot user (bot score <*0.5*) compared to a bot user (bot score >*0.5*). Table 1 presents the results of the model involving hurricane related tweets. All the variables are significant at 90 percent significance level. It is found that efficiency increases with initial number of followers *(0.002),* bot score of bot users *(11.622),* initial number of followee of non-bot users *(-0.004+0.007=0.003)*; while it decreases with total activity (-0.039), initial number of followees of bot users (*-0.004*), bot score of non-bot users *(11.622 - 14.419=-2.797).* A negative coefficient (*-36.754*) for activity entropy implies that entropy values have negative correlation with efficiency and users having a predictable activity pattern (lower entropy values) have higher efficiency values. The model estimated over Sandy specific tweets show similar results (see

Table S3) except that total activity is not statistically significant in this case and bot score *(1.053 for bot user, 1.053+0.521=1.573 for non-bot user)* and activity entropy *(0.704)* are positively correlated with efficiency.

From the regression analysis, we find how different user features influence a user's efficiency. However, while considering overall efficiency as an outcome, a minor change in efficiency does not provide any significant information about the user's performance in gaining attention. Thus, we have categorized efficiency into five classes: negative (efficiency<0), zero (efficiency=0), low (0<efficiency<=5), moderate (5<efficiency<=10) and high (efficiency>10) as shown in Figure 5. The outcome variable thus turned into an ordered categorical variable. In the hurricane related sample about 6%, 56%, 32%, 3%, and 3% of the users fall within the negative, zero, low, moderate and high efficiency category, respectively. We have estimated an ordered logit model using the same 582605 observations (see Table 1). All the parameters shown in the results are statistically significant at 99% significance level.

TABLE 1 Model Results for hurricane related tweets

Linear Regression Model			**Ordered Logit Model**		
Explanatory variable	Parameter Estimate	t statistic	Parameter Estimate	Odds Ratio Estimates	
				Point Estimate	95% Wald Confidence Limits
Intercept 5	na		-5.0145		
Intercept 4	na		-4.2164		
Intercept 3	na		-1.1988		
Intercept 2	na		2.2357		
Constant	0.758	1.712	Na		

Total activity	-0.039	-6.767	-0.0110	0.989	0.989	0.989
Followee add	0.098	77.845	0.00926	1.009	1.009	1.009
Initial follower	0.002	698.21	1.000	1.000	1.000	1.000
Initial followee	-0.004	-53.3	1.000	1.000	1.000	1.000
Active days	na		0.2951	1.343	1.340	1.347
Activity entropy	-36.754	-15.493	-1.9523	0.142	0.128	0.157
Bot score	11.622	9.793	-0.3974	0.672	0.648	0.697
Bot_Score*NonBotUser	-14.419	-10.041	Na			
Initial_followee*NonBotUser	0.007	37.423	Na			
Number of observations	582,605		582,605			
Adjusted R_2	0.47					
AIC			1244359.6			

Note: na= not applicable, any variable included in one mode but not included in the other one is because we did not find it significant.

For the interpretation of this result, a negative coefficient of total activity *(-0.0110)* indicates that if all other variables in the model remain constant, for an increase in the number of tweets, a user is more likely to be in a lower level of efficiency. Similar to the results from the regression model, users with a predictable tweeting pattern *(-1.9523),* (i.e., smaller entropy value) are more likely to be in a higher efficiency category. Moreover, we find that users with higher number of active days *(0.2951)* and followee added *(0.00926)* are more likely to be in a higher category of efficiency. In contrast, users with higher total activity *(-0.0110)* and bot score *(-0.3974)* are less likely to be in higher category efficiency. An ordered Logit model estimated over the Sandy specific tweets shows similar association except that total activity is not statistically significant and activity entropy is positively correlated with efficiency (see Table S3). A higher number of initial follower *(1.00)* or initial followee *(1.00)* does not result in higher or lower efficiency category for both hurricane related and Sandy specific tweets.

6. Discussion

During a hurricane, affected people turns to social media for updates and assistance from individuals and organizations. In addition, emergency management agencies use social media to organize and direct resources where needed. These interactions among social media users and agencies create dynamics in communication that varies over different phases of a disaster. We explore the dynamics of crisis communication using hurricane Sandy Twitter data. We define an efficiency metric as the ratio between attention received (gaining new followers) and activity (posting tweets/retweets) in a given time interval. From the activity-attention analysis, we find that higher (>100) followee addition causes higher attention (follower received) in all three (pre-disaster, during disaster and post-disaster) phases of a hurricane. High reciprocity (e.g., mutual follower-followee relationship) among social media users might be responsible for such attention gain. For instance, during a hurricane, individuals may follow each other to get updates and to get connected for taking common decisions (e.g., evacuation) (Sadri, Hasan, & Ukkusuri, 2017).

In social media platforms, all users may not get equal attention for their activities. Based on the efficiency values, we categorize five (negative, zero, low, moderate, high) types of Twitter users. Majority of the users in our dataset are in the first three categories (negative, zero, low). A negative (overall efficiency <0) value of the efficiency metric means that the user lost followers during the analysis period. This also confirms previous studies (Myers & Leskovec, 2014) that information diffusion creates sudden burst in creating new connections or deleting existing connections. The zero (overall efficiency =0) efficiency category represents no change in attention dynamic despite their activities (tweeting/retweeting). The low (0<overall efficiency≤5) and moderate (5<overall efficiency≤10) efficiency categories, however, do not show any trend in the efficiency dynamics during different phases of the hurricane. For high efficiency (overall efficiency>10) category, the dynamics of efficiency metric shows that efficiency is the highest just after the official declaration of a hurricane (see Figure 5), which indicates the promptness of public reaction and participation in information sharing or consumption during an upcoming disaster. After the highest peak, the efficiency of the highly efficient users has a decreasing trend with a peak on landfall day and some peaks in the recovery period (see Figure 5). This shows that highly efficient users have consistently gained attention from a pre-disaster period to a post-disaster period with the most attention gained right after the hurricane declaration.

Individuals are likely to follow those users who post relevant news and updates. Investigating the users who made Sandy specific tweets (see Table S2), we find that majority of the top 5 efficient users are storm related organizations, that post relevant news and information about the hurricane. This confirms our hypothesis that individuals follow users who publishes relevant updates about the hurricane. In case of unfiltered data, top 5 users (see Table S1) are mostly political users (political campaign of US presidential election 2012 was going on at that time). This finding also indicates the importance of appropriate filtering to retrieve more relevant information from social media data.

We have estimated models to determine the key attributes that make a user more efficient in gaining attention in an emergency situation. Our estimated models show that higher activity frequency is likely to decrease efficiency, which is similar to the findings of this study (Vaca et al., 2014) on microblogging data. We have also found that increasing followee is likely to increase the efficiency of gaining attention, similar to the finding of this study (Cha, Haddai, Benevenuto, & Gummadi, 2010) that influential users need to maintain constant engagement. Initial number of followers has a positive (very small coefficient) effect in the regression model but does not result in higher or lower efficiency category in the logit model (odds ratio =1), indicating that a higher number of initial followers may increase efficiency but not enough to change the efficiency category. A previous study (Cha et al., 2010) also found that a high number of initial followers may represent popularity but not in terms of user engagement.

Interestingly, in the regression model initial number of followee has a positive coefficient for a non bot user but a negative coefficient for a bot user. It indicates that for an individual user (operated by a human) a higher number of initial followee helps gain attention. But for a bot user (e.g., an organizational account), a higher number of initial followee has a negative effect in gaining efficiency. However, in terms of efficiency category, initial followee does not result in higher or lower efficiency category in the logit model (odds ratio=1).

From the estimated models, bot score shows mixed results for hurricane related tweets and Sandy specific tweets. For Sandy specific model, bot score has a positive coefficient for gaining efficiency, which indicates the potential of using a bot account to generate, response, and share hurricane specific content during a disaster to gain higher efficiency. However, for hurricane related tweets, bot score has a positive coefficient for bot users, but a negative coefficient for non-bot users. It means that an account which is already acting like a bot is likely to have higher efficiency by increasing bot-like hurricane related activity. In contrast, a non-bot account is likely to have lower efficiency for increasing bot-like activity related to hurricane.

In addition, for the Sandy specific model, active days and activity entropy have negative and positive coefficients, respectively. This implies that less active days with irregular but specific topic focused posts are likely to help gaining efficiency. On the other hand, from the hurricane related model, we see that active days and activity entropy have positive and negative coefficients, respectively. This means that more active days with regular hurricane related posts are likely to help gaining efficiency. As hurricane related contents contain a lot of topics, regular and predictive activity pattern related to hurricane can help maintaining constant engagement, an important factor for influence as found in other studies (Cha et al., 2010) .

In summary, the results of this study indicate the nature and dynamics of efficiency of attention gain from a pre-disaster period to a post-disaster period. These findings also shed light on the important factors to gain efficiency—adding valuable insights to social-mediated crisis communication theory and literature. Practitioners such as emergency managers and public-relation managers can also find implications towards adopting better communication strategies. A detailed discussion on the contribution and practical implications of our results are given in the following sections.

6.1 Contributions to theory/literature

Social media usage during disasters has changed the way how organizations manage crisis communication. People are increasingly using social media during a crisis and thus crisis communication professionals need to understand—how to optimize activity and engagement to maximize the spreading of situational awareness messages. In addition, monitoring and responding to the content of influential social media creators are important to efficiently communicate with the targeted people, correcting any misinformation and rumors. Thus, interconnection among the organization, influential social media creators and the followers are a very important component in the social-mediated crisis communication (see Figure 1). To date, in the context of blog mediated crisis communication model, few studies (Liu et al., 2012) investigated the metrics to evaluate the influence of a blog. These metrics include blog characteristics, blogger credibility, interconnectivity, frequency of updates, consistency with the organization's crisis response, etc. However, these studies mainly focus on typical blogsphere content and are not tuned for social media platforms such as Twitter, Facebook etc. Recent studies on social media influence mainly focus on the content not on the users. In our study, we have proposed a simple attention-based efficiency measure and have considered features such as activity frequency, activity pattern, active days, number of friends and followers, followee add, bot score etc. to identify efficient users. Here, the features, activity frequency and number of friends/followers in social media contexts can be considered similar to frequency of updates and inter-connectivity in the context of blogs. Many organizations use Twitter bot for tweeting, reposting, and messaging now a days. We adopt the bot score feature in our study to capture the effect of automation level in gaining attention efficiency. This study thus provides an empirical evidence on the dynamics of efficiency from a pre-disaster to a post-disaster period. It also provides insights on the influence of different efficiency factors that can help better understand social mediated crisis communication.

6.2 Implications for Practice

This study also provides valuable insights for practitioners such as crisis communication and public relation managers. From efficiency dynamics, we can understand when influential users get more attention (see figure 5). Crisis managers should monitor social media more during these times to ensure dissemination of authentic and most recent news. The findings from the contributing factors analysis provide insight about activity optimization strategies to get more attention during crisis communication. Some of the implications of this study are given below

Activity Frequency: Influential individuals and relevant organizations post (tweet) or repost (retweet) about different situational awareness during a disaster. These users need to optimize their posting frequency as increased activity frequency has a negative effect in gaining attention efficiency (see Table 1 and Table s3). Crisis managers thus need to maintain their activity only to post meaningful updates rather than posting redundant and unimportant messages.

Followee add: A crisis communication manager aims to maximize the information spreading to as many new users as possible. The crisis manager also needs to monitor social media content to prevent misinformation/rumors. We have found that followee addition has a positive effect in getting attention efficiency. So, a crisis manager can follow other users during a crisis, especially the influential users, who are likely to help gain more attention to emergency messages.

Initial follower: A highly connected user can reach to many users and thus can disseminate information to a larger audience. We have found (see Table 1 and Table s3) that a high number of initial followers are likely to help to be efficient in gaining attention from the users who are not a follower yet. An organizational account with a greater number of followers is more likely to spread information faster. Similarly, a crisis communication manager can monitor users with high connectivity (followers) as these users are likely to be the influential users during a crisis.

Bot Score: Social bots are programmed to create posts in social media platforms autonomously. Approximately 15 percent of Twitter accounts are bots rather than humans (Newberg, 2017). Although there are evidences that social bots have been used to manipulate public opinion, creating false impression and misinformation especially during elections, a social bot has many good applications such as disseminating weather news, crisis information, evacuation alerts etc. During emergencies a social bot can disseminate information about situational awareness and respond to the users seeking help or information. During emergencies, social media managers handle unprecedented amount of activities, and because of lack of trained personnel only few people work hard to manage the activities during such events (Oyeniyi, 2017). The main task of a social media crisis manager is to disseminate information by posting in social media platforms (Oyeniyi, 2017). Social bots can reduce the workload during crisis by automating some of the routine tasks so that a crisis manager can monitor social media posts to reduce misinformation/rumors. According to our findings, bot score has a positive effect in gaining efficiency for bot users (bot score > 0.5) (see Table 1 and Table S3)—suggesting that users having bot characteristics are more likely to be efficient in gaining attention. If a user account/organization has a high bot score or already using automation, using social bots to generate and disseminate messages are likely to help gain more attention from new users (not follower yet). However, for a non-bot user (bot score < 0.5), bot score has a mixed effect (see Table 1 and Table S3). Thus, if an emergency manager uses a non-bot account to disseminate information, adopting a social bot may not help gain higher attention efficiency (see Table S3).

7. Conclusions

In this study, we have analyzed twitter posts related to hurricane Sandy to understand the effectiveness of social media-based communication during disasters. Effective crisis communication can ensure faster information dissemination to vulnerable communities who need timely information about disaster preparedness, evacuation warning, and recovery operations. To measure the effectiveness of a social media user in communicating information or awareness, we have estimated the efficiency of gaining attention within a specific time period as the ratio of follower gained over tweet frequency within the same time period. We consider that new follower gained represents attention received and tweet frequency represents activities made. As our data contains tweets from both pre and post-disaster periods, a comparison of user efficiencies in gaining attention among these periods has been possible.

Analyzing daily efficiencies in gaining attention, we have found that users had higher efficiency during the critical periods in hurricane Sandy such as declaration and landfall days. This indicates the potential of social media-based crisis communication since higher attention to related information may help in providing situational awareness to vulnerable population. It might be the case that during Sandy some users' efficiency became abnormally high because a high number of users started following them for hurricane related updates. These social media users could become one of the major sources of information for spreading hurricane awareness during future hurricanes.

During a disaster, general social media users seek information from other users for a timely update. When sharing information, some users gain more attention than the others. Thus, it's critical to understand what user features influence the process of gaining attention. For understanding the contributing features, we have estimated a regression and an ordered logit model considering overall efficiency and efficiency category, respectively as a dependent variable. We have found that higher activities relating to hurricanes are not necessarily associated with higher efficiencies. However, users with predictable tweeting patterns have gained higher efficiency values. We have also found that a higher bot score (typically associated with an organizational account) results in lower efficiencies. We have observed some differences on the effect of few user attributes on efficiency values for models estimated over general hurricane related tweets and Sandy specific tweets. User efficiencies in gaining attention for a crisis event are directly related with information spreading capacity of a system. A better understanding of the factors will provide insights on crisis communication both at organizational and individual levels. These insights will also help emergency agencies when using social media as a disaster communication tool.

Thus, our findings have significant importance in social media communication specially in disaster communication. For attaining high efficiency in spreading disaster related information, concerned organizational or personal accounts can plan their activity considering the factors which will maximize the chance to attain higher attention from the targeted population in social media. Also, prior to a major disaster event concerned authorities can select some efficient social media users for disseminating information about situational awareness; a model based on user features could find the efficient users for this task.

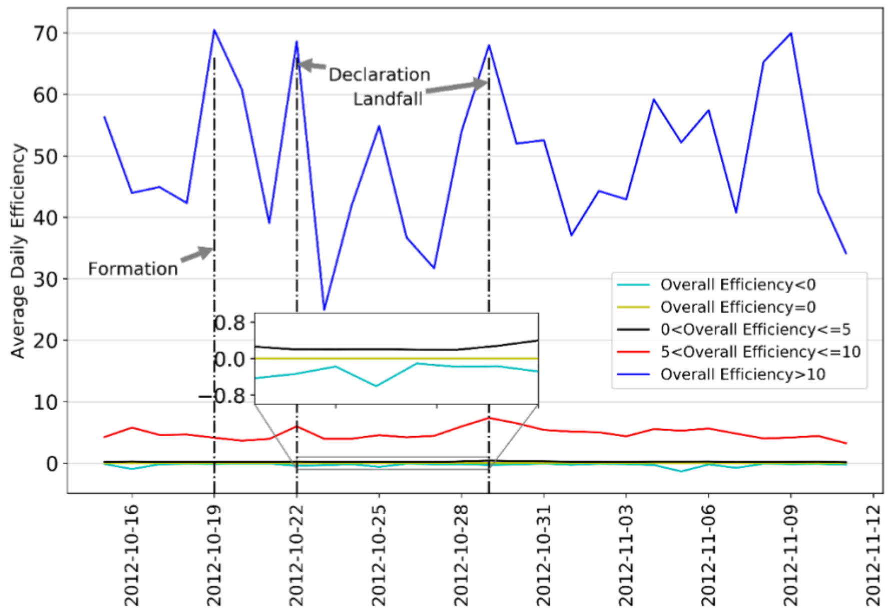
7.1 Limitations and Future Research Directions

Our study has several limitations which can be improved in future. For example, bot scores were not collected during the period of hurricane Sandy; rather we have collected them at the time of our analysis. Bot scores could be different during hurricane Sandy than in the present. We assume that all the tweets analyzed here were related to hurricane Sandy. User efficiency considering specific topics (e.g., evacuation) of a tweet should be analyzed in the future. We have not checked the bot scores of the newly gained followers. The newly gained followers may be social bots instead of human accounts. Studies are needed to understand – how attention to situational awareness messages actually turn into action. For example, we need to investigate if information related to evacuation order is disseminated to affected people, how they respond to such orders. Finally, our results are based on only one event; data from multiple events should be analyzed to better understand the influence of social media during a crisis event.

Acknowledgment

This work was supported by the U.S. Department of Transportation's University Transportation Centers Program and the U.S. National Science Foundation (Grant IDs CMMI #1832578 and #1917019). The authors are solely responsible for the facts and accuracy of the information presented in the paper.

Supporting information



A

Fig S1. Average Daily Efficiency of the users categorized by efficiency (for hurricane related tweets). X-axis shows the time in date and Y-axis shows the average daily efficiency. Users are categorized by their overall efficiency values.

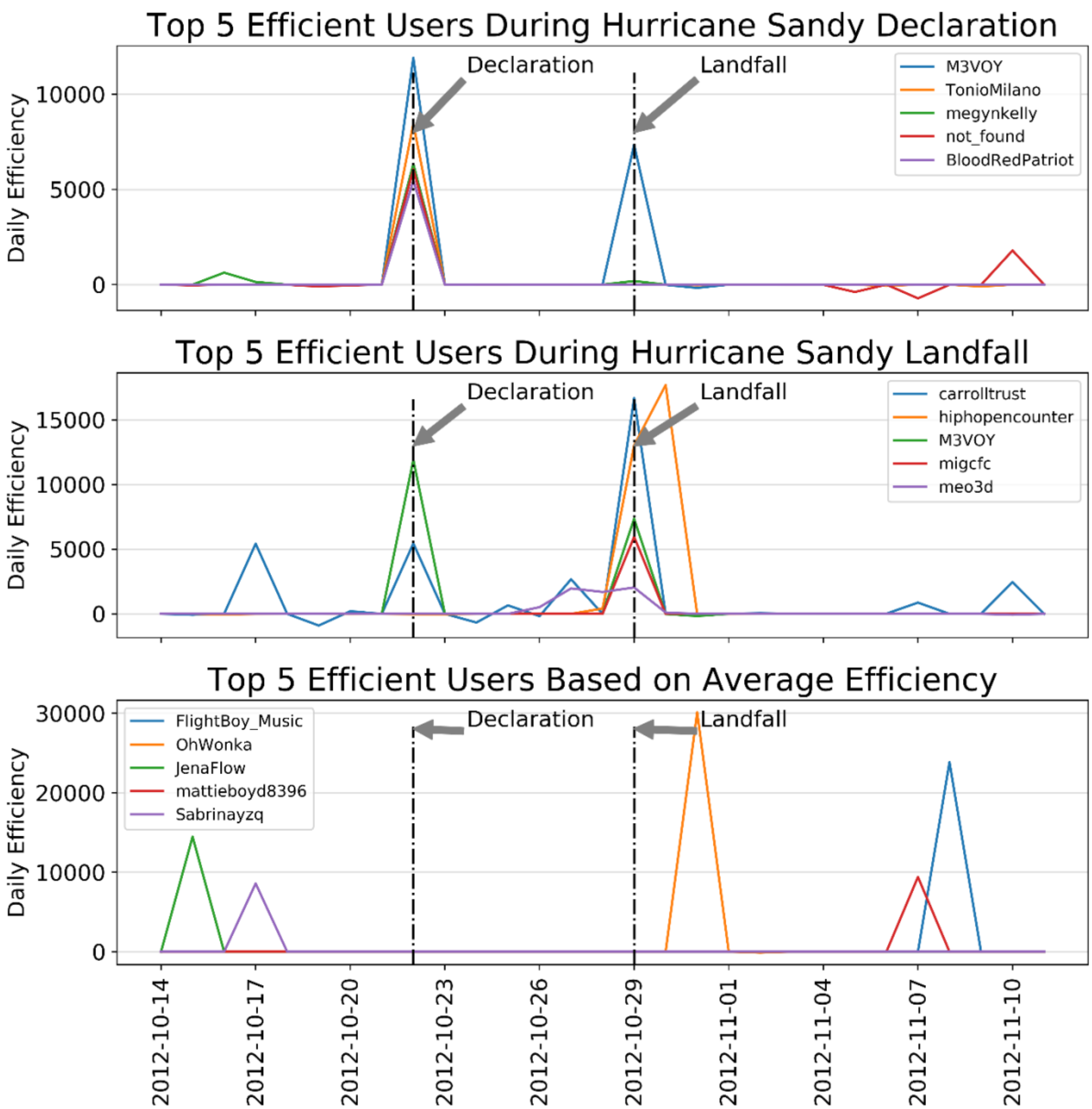


Fig S2. Top 5 Efficient Users during Hurricane Declaration, Landfall and Overall (for hurricane related tweets).

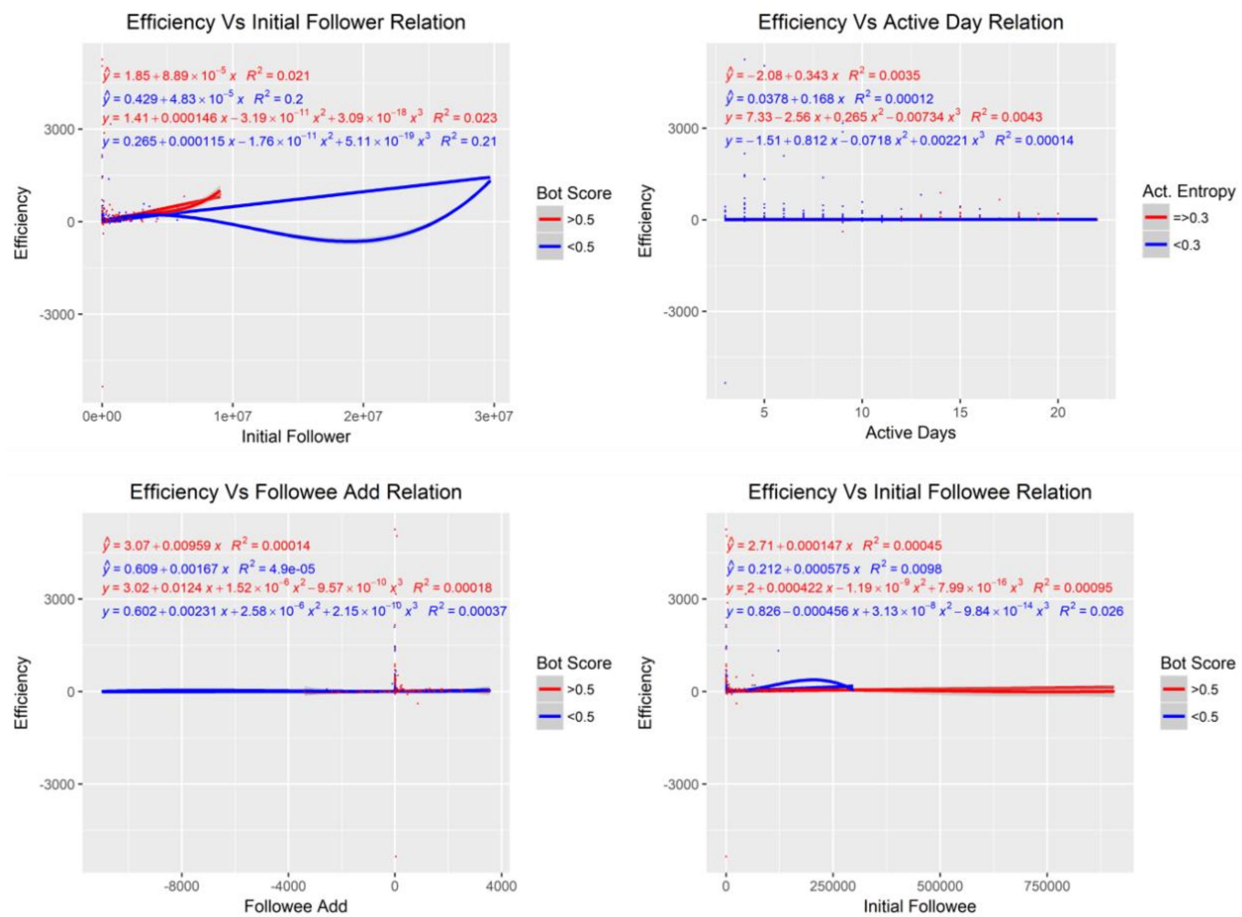


Fig S3. Relationship between Efficiency and Different User Attributes (for sandy specific tweets)

Table S1. Top 5 efficient users with user type and bot score during hurricane declaration and hurricane landfall. This analysis is based on the unfiltered (hurricane related tweets) data.

Hurricane Declaration			**Hurricane Landfall**		
Screen Name	User Type	Bot Score	Screen Name	User Type	Bot Score
@M3VOY	Radio Operator	0.62	@carrolltrust	Organization	0.43

@TonioMilano	Personal	0.48	@hiphopencounter	Organization	0.76
@megynkelly	Anchor at NBC News	0.36	@migcfc	Personal	0.06
@BloodRedPatriot	Organization	0.73	@trio	Organization	0.55

Table S2. Top 5 efficient users with user type and bot score during hurricane declaration and landfall. This analysis is based on the data which have the word sandy in the tweet.

Hurricane Declaration			**Hurricane Landfall**		
Screen Name	User Type	Bot Score	Screen Name	User Type	Bot Score
Not Found	NA	NA	@NHC_Atlantic	Organization	0.43
@breakingstorm	Organization	0.66	@cnnbrk	Organization	0.51
@NHC_Atlantic	Organization	0.43	@Jimcantore	Broadcast Meteorologist	0.37
@Jimcantore	Broadcast Meteorologist	0.37	@BreakingNews	Organization	0.60
@kkstormcenter	Weather reporter	0.71	@breakingstorm	Organization	0.66

Note: Not found indicates that the user made tweets during Sandy but its screen name was not found when searched during our analysis.

TABLE S3 Model Results for Sandy specific tweets

Linear Regression Model			Ordered Logit Model			
Explanatory variable	Parameter Estimate	t statistic	Parameter Estimate	Odds Ratio Estimates		
				Point Estimate	95% Wald Confidence Limits	
Intercept 5	na		-6.6001			
Intercept 4	na		-3.7514			
Intercept 3	na		4.6505			
Intercept 2	na		5.3600			
Constant	-.073	-.453	na			
Total activity	-0.001	-1.265	na	na	na	na
Followee add	0.003	5.411	0.00359	1.004	1.003	1.004
Initial follower	0.0001	231.54	0.000035	1.00	1.00	1.00
Initial followee	-6.85E-5	-6.219	-2.71E-06	1.000	1.000	1.000
Active days	na		-0.1312	0.877	0.852	0.903
Activity entropy	.704	1.940	1.2244	3.402	1.726	6.708
Bot score	1.053	3.735	2.115	8.289	5.227	13.145
Bot_Score*NonBotUser	.521	1.655	na			
Initial_followee*NonBotUser	-5.27E-5	-1.794	na			
Number of observations	15,792		15,792			
Adjusted R^2	0.77					
AIC			7513.390			

Note: na= not applicable, any variable included in one mode but not included in the other one is because we did not find it significant.

References

Aguirre, B. E., Wenger, D., & Vigo, G. (1998). A Test of the Emergent Norm Theory of Collective Behavior. *Sociological Forum*, *13*(2), 301–320, Kluwer Academic Publishers-Plenum Publish.

Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, *34*(7), 1177–1190.

Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology & Marketing*, *32*(1), 15–27.

Asur, S., Huberman, B. A., Szabo, G., & Wang, C. (2011). Trends in Social Media: Persistence and Decay. *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11)*, 434–437. Retrieved from http://arxiv.org/abs/1102.1402

Bharosa, N., Lee, J., & Janssen, M. (2010). Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers*, *12*(1), 49–65.

Blake, E. S., Kimberlain, T. B., Berg, R. J., Cangia, Losi, J. P., & Beven II, J. L. (2013). Tropical cyclone report Hurricane Sandy (AL182012) 22 – 29 October 2012. *National Weather Service, National Hurricane Center*, (October 2012), 1–157.

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*, 10–17. Retrieved from http://www.icwsm.org/2010/

Cha, M., Mislove, A., & Gummadi, K. P. (2009). A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. *Proceedings of the 18th International Conference on World Wide Web*, 721–730. Retrieved from http://doi.acm.org/10.1145/1526709.1526806

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers and Graphics (Pergamon)*, *38*(1), 51–60. Retrieved from http://dx.doi.org/10.1016/j.cag.2013.10.008

Chen, Y., Frei, A., & Mahmassani, H. (2014). From Personal Attitudes to Public Opinion : Information Diffusion in Social Networks towards Sustainable Transportation. *Transportation Research Record: Journal of the Transportation Research Board*, *2430*, 28–37.

Chen, Y., Mahmassani, H. S., & Frei, A. (2017). Incorporating social media in travel and activity choice models: conceptual framework and exploratory analysis. *International Journal of Urban Sciences*, *0*(0), 1–21. https://doi.org/10.1080/12265934.2017.1331749

Comfort, L. K., Ko, K., & Zagorecki, A. (2004). Coordination in Rapidly Evolving Disaster Response Systems. *American Behavioral Scientist*, *48*(3), 295–313. Retrieved from http://journals.sagepub.com/doi/10.1177/0002764204268987

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot: A System to Evaluate Social Bots. *ArXiv*, : *1602.009*. Retrieved from http://arxiv.org/abs/1602.00975

Demuth, J. L., Morss, R. E., Palen, L., Anderson, K. M., Anderson, J., Kogan, M., … others. (2018). “Sometimes da# beachlife ain’t always da wave”: Understanding People’s Evolving Hurricane Risk Communication, Risk Assessments, and Responses Using Twitter Narratives. *Weather, Climate, and*

Society, *10*(3), 537–560.

Derr, B. (2013). Ordinal Response Modeling with the LOGISTIC Procedure. *SAS Global Forum*, 1–20.

Dong, R., Li, L., Zhang, Q., & Cai, G. (2018). Information Diffusion on Social Media During Natural Disasters. *IEEE Transactions on Computational Social Systems*, *5*(1), 265–276.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2014). The Rise of Social Bots. *ArXiv Preprint ArXiv:1407.5225*, (grant 220020274), 1–11. Retrieved from http://arxiv.org/abs/1407.5225

Figueiredo, F., Benevenuto, F., Almeida, J. M., Fabr, F. F., & Almeida, B. J. M. (2011). The tube over time: characterizing popularity growth of youtube videos. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining - WSDM '11*, 745–754. Retrieved from http://portal.acm.org/citation.cfm?doid=1935826.1935925

Hasan, S., Mesa-Arango, R., & Ukkusuri, S. (2013). A random-parameter hazard-based model to understand household evacuation timing behavior. *Transportation Research Part C: Emerging Technologies*, *27*(January), 108–116. Retrieved from http://dx.doi.org/10.1016/j.trc.2011.06.005

Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, *44*, 363–381. Retrieved from http://dx.doi.org/10.1016/j.trc.2014.04.003

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing - UrbComp '13*. Retrieved from http://dl.acm.org/citation.cfm?doid=2505821.2505823

Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in Twitter. *Proceedings of the 20th International Conference on World Wide Web (WWW)*, 57–58. Retrieved from http://portal.acm.org/citation.cfm?doid=1963192.1963222

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168–177).

Huang, C.-M., Chan, E., & Hyder, A. a. (2010). Web 2 . 0 and Internet Social Networking : A New tool for Disaster Management ? - Lessons from Taiwan. *BMC Medical Informatics and Decision Making*, *10*(1), 57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20925944

Huang, Q., & Xiao, Y. (2015). Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. *ISPRS International Journal of Geo-Information*, *4*(3), 1549–1568. Retrieved from http://www.mdpi.com/2220-9964/4/3/1549/htm

Jin, Y., & Liu, B. F. (2010). The blog-mediated crisis communication model: Recommendations for responding to influential external blogs. *Journal of Public Relations Research*, *22*(4), 429–455.

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: past, present and future. *Information Systems Frontiers*, 1–28.

Kass-Hout, T. A., & Alhinnawi, H. (2013). Social media in public health. *British Medical Bulletin*, *108*(1), 5–24. https://doi.org/10.1093/bmb/ldt028

Kaufman, S., C. Qing, N. Levenson, and M. H. (2012). Transportation during and after Hurricane Sandy.

Kim, J., Bae, J., & Hastak, M. (2018). Emergency information diffusion on online social media during storm Cindy in US. *International Journal of Information Management*, *40*, 153–165.

Kim, J., & Hastak, M. (2018). Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management*, *38*(1), 86–96.

Kim, W.-S., Yoon, Y.-Z., Bae, J.-H., & Soh, K.-S. (2005). Nonlinear characteristics of heart rate time series: influence of three recumbent positions in patients with mild or severe coronary artery disease. *Physiological Measurement*, *26*(4), 517–529. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15886445

Kryvasheyeu, Y., & Chen, H. (2015). Performance of Social Network Sensors During Hurricane Sandy. *PLoS One*, *10*(2), e0117288. Retrieved from http://arxiv.org/abs/1402.2482

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. Van, Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances 2.3 (2016): E1500779.*

Lehmann, J., Gonçalves, B., Ramasco, J. J., & Cattuto, C. (2012). Dynamical Classes of Collective Attention in Twitter. *Proceedings of the 21st International Conference on World Wide Web. ACM*. Retrieved from http://arxiv.org/abs/1111.1896

Leung, D., Law, R., van Hoof, H., & Buhalis, D. (2013). Social Media in Tourism and Hospitality: A Literature Review. *Journal of Travel and Tourism Marketing*, *30*(1–2), 3–22. https://doi.org/10.1080/10548408.2013.750919

Lindsay, B. R. (2011). Social Media and Disasters: Current Uses, Future Options and Policy Considerations. Retrieved from http://fas.org/sgp/crs/homesec/R41987.pdf

Liu, B. F., Jin, Y., Briones, R., & Kuch, B. (2012). Managing turbulence in the blogosphere: Evaluating the blog-mediated crisis communication model with the American Red Cross. *Journal of Public Relations Research*, *24*(4), 353–370.

MacMillan, D. (2017). In Irma, Emergency Responders' New Tools: Twitter and Facebook. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/for-hurricane-irma-information-officials-post-on-social-media-1505149661

Martín, Yago, Zhenlong Li, and S. L. C. (2017). Leveraging Twitter to gauge evacuation compliance: spatiotemporal analysis of Hurricane Matthew. *PLoS One*, (July). https://doi.org/10.1371/journal.pone.0181701

Mathioudakis, M., Koudas, N., & Marbach, P. (2010). Early online identification of attention gathering items in social media. *Proceedings of the Third ACM International Conference on Web Search and Data Mining - WSDM '10*, 301. Retrieved from http://portal.acm.org/citation.cfm?doid=1718487.1718525

Mileti, D. S., & Fitzpatrick, C. (1992). The causal sequence of risk communication in the Parkfield earthquake prediction experiment. *Risk Analysis*, *12*(3), 393–400.

Misirlis, N., & Vlachopoulou, M. (2018). Social media metrics and analytics in marketing--S3M: A mapping literature review. *International Journal of Information Management*, *38*(1), 270–276.

Munich RE. (2018, January 4). Natural catastrophe review: Series of hurricanes makes 2017 year of highest insured losses ever. Retrieved January 15, 2018, from https://www.munichre.com/en/media-relations/publications/press-releases/2018/2018-01-04-press-release/index.html

Murray-Tuite, P., & Wolshon, B. (2013). Evacuation transportation modeling: An overview of research, development, and practice. *Transportation Research Part C: Emerging Technologies*, *27*, 25–45. Retrieved from http://dx.doi.org/10.1016/j.trc.2012.11.005

Myers, S. A., & Leskovec, J. (2014). The bursty dynamics of the twitter information network. In *Proceedings of the 23rd international conference on World wide web* (pp. 913–924).

NELSON, B. (2013). A calculated risk. *Nature*, *495*, 271–273.

Newberg, M. (2017). As many as 48 million Twitter accounts aren't people, says study. Retrieved from https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html

Ngai, E. W. T., Tao, S. S. C., & Moon, K. K. L. (2015). Social media research: Theories, constructs, and conceptual frameworks. *International Journal of Information Management*, *35*(1), 33–44.

Ni, M., He, Q., & Gao, J. (2017). Forecasting the Subway Passenger Flow under Event Occurrences with Social Media. *IEEE Transactions on Intelligent Transportation Systems*, *18*(6), 1623–1632.

https://doi.org/10.1109/TITS.2016.2611644

NOAA National Centers for Environmental Information (NCEI) U.S. (2018). Billion-Dollar Weather and Climate Disasters. Retrieved January 15, 2018, from https://www.ncdc.noaa.gov/billions/

Oyeniyi, D. (2017). How Hurricane Harvey Changed Social Media Disaster Relief. *TexasMonthly*. Retrieved from https://www.texasmonthly.com/the-daily-post/how-social-media-managers-responded-to-hurricane-harvey/

Palen, L., Vieweg, S., Liu, S. B., & Hughes, A. L. (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event. *Social Science Computer Review*, *27*(4), 467–480. Retrieved from http://ssc.sagepub.com/cgi/doi/10.1177/0894439309332302

Paton, D. (2008). Risk communication and natural hazard mitigation: how trust influences its effectiveness. *International Journal of Global Environmental Issues*, *8*(1–2), 2–16.

Pincus, S., & Kalman, R. E. (2004). Irregularity, volatility, risk, and financial market time series. *Proceedings of the National Academy of Sciences*, *101*(38), 13709–13714. Retrieved from http://www.pnas.org/cgi/doi/10.1073/pnas.0405168101

PINCUS, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences 88.6 (1991): 2297-2301.*

Pincus, SM, & Goldberger, A. (1994). Physiological time-series analysis: What does regularity quantify? *The American Journal of Physiology*, *266*(4 Pt 2), H1643–H1656.

Quercia, D., Ellis, J., Capra, L., & Crowcroft, J. (2011). In the mood being influential on twitter mood. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, 307–314.

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, *75*, 197–211. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0968090X16302625

Reuter, C., & Kaufhold, M. (2017). Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics. *Wiley Online Library*. https://doi.org/https://doi.org/10.1111/1468-5973.12196

Romero, D. M., & Huberman, B. A. (2011). Influence and Passivity in Social Media Background on Twitter. *WWW '11 Proceedings of the 20th International Conference Companion on World Wide Web*.

Sadri, A. M., Hasan, S., & Ukkusuri, S. V. (2017). Joint Inference of User Community and Interest Patterns in Social Interaction Networks. *ArXiv Preprint ArXiv:1704.01706*. Retrieved from http://arxiv.org/abs/1704.01706

Sadri, A. M., Hasan, S., Ukkusuri, S. V., & Cebrian, M. (2017). Understanding Information Spreading in Social Media during Hurricane Sandy: User Activity and Network Properties. *ArXiv Preprint ArXiv:1706.03019*. Retrieved from http://arxiv.org/abs/1706.03019

Sadri, A. M., Hasan, S., Ukkusuri, S. V., & Lopez, J. E. S. (2017). Analyzing Social Interaction Networks from Twitter for Planned Special Events. *ArXiv Preprint ArXiv:1704.02489*. Retrieved from http://arxiv.org/abs/1704.02489

Sadri, A. M., Ukkusuri, S. V., & Murray-Tuite, P. (2013). A random parameter ordered probit model to understand the mobilization time during hurricane evacuation. *Transportation Research Part C: Emerging Technologies*, *32*(November 2015), 21–30. Retrieved from http://dx.doi.org/10.1016/j.trc.2013.03.009

Sadri, A. M., Ukkusuri, S. V., Murray-Tuite, P., & Gladwin, H. (2014). Analysis of hurricane evacuee mode choice behavior. *Transportation Research Part C: Emerging Technologies*, *48*(November 2015), 37–46. Retrieved from http://dx.doi.org/10.1016/j.trc.2014.08.008

Shiau, W.-L., Dwivedi, Y. K., & Lai, H.-H. (2018). Examining the core knowledge on facebook. *International Journal of Information Management*, *43*, 52–63.

Shiau, W.-L., Dwivedi, Y. K., & Yang, H. S. (2017). Co-citation and cluster analyses of extant literature on social networks. *International Journal of Information Management*, *37*(5), 390–399.

Shklovski, I., Palen, L., & Sutton, J. (2008). Finding Community Through Information and Communication Technology During Disaster Events. *2008 ACM Conference on Computer Supported Cooperative Work*, 127–136.

Simon, T., Goldberg, A., & Adini, B. (2015). Socializing in emergencies - A review of the use of social media in emergency situations. *International Journal of Information Management*, *35*(5), 609–619. Retrieved from http://dx.doi.org/10.1016/j.ijinfomgt.2015.07.001

Srinivasan, V., Eswaran, C., & Sriraam, N. (2007). Approximate Entropy-Based Epileptic EEG Detection Using Artificial Neural Networks, *11*(3), 288–295.

Stai, E., Milaiou, E., Karyotis, V., & Papavassiliou, S. (2018). Temporal Dynamics of Information Diffusion in Twitter: Modeling and Experimentation. *IEEE Transactions on Computational Social Systems*, *5*(1), 256–264. https://doi.org/10.1109/TCSS.2017.2784184

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics--Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168.

The World Bank. (2016). Natural Disasters Force 26 Million People into Poverty and Cost $520bn in Losses Every Year, New World Bank Analysis Finds. Retrieved June 15, 2017, from http://www.worldbank.org/en/news/press-release/2016/11/14/natural-disasters-force-26-million-people-into-poverty-and-cost-520bn-in-losses-every-year-new-world-bank-analysis-finds

Torres-reyna, O. (2012). Getting Started in Logit and Ordered Logit Regression Logit model. *Princeton University , URL Http://Dss. Princeton. Edu/Training/Logit. Pdf*. Retrieved from URL

Ukkusuri, S., Zhan, X., Sadri, A., & Ye, Q. (2014). Use of Social Media Data to Explore Crisis Informatics. *Transportation Research Record: Journal of the Transportation Research Board*, *2459*, 110–118.

Vaca, C., Aiello, L. M., Jaimes, A., & Milano, P. (2014). Modeling Dynamics of Attention in Social Media with User Efficiency. *EPJ Data Science 3.1 (2014): 5.*

Washington, S.P., Karlaftis, M.G. and Mannering, F. (2010). Statistical and Econometric Methods for Transportation Data Analysis. *CRC Press*.

Wu, F., Huberman, B. A., Adamic, L. A., & Tyler, J. R. (2004). Information flow in social groups. *Physica A*, *337*(1), 327–335.

Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management*, *31*(1), 6–13. Retrieved from http://dx.doi.org/10.1016/j.ijinfomgt.2010.10.001

Zhang, Z., He, Q., Gao, J., & Ni, M. (2018). A deep learning approach for detecting traffic accidents from social media data. *Transportation Research Part C: Emerging Technologies*, *86*, 580–596. https://doi.org/10.1016/j.trc.2017.11.027