Exploring network properties of social media interactions and activities during Hurricane Sandy

Arif Mohaimin Sadri [a,*], Samiul Hasan [b], Satish V. Ukkusuri [c], Manuel Cebrian [d]

[a] Florida International University, United States of America
[b] University of Central Florida, United States of America
[c] Purdue University, United States of America
[d] Max Planck Institute for Human Development, Germany

ARTICLE INFO

ABSTRACT

In this study, we analyze Twitter data to understand information spreading activities of social media users during Hurricane Sandy. We create multiple subgraphs of Twitter users based on activity levels and analyze such network properties. We observe that user information sharing activity follows a power-law distribution suggesting the existence of few highly active nodes in disseminating information compared to many other nodes. We also observe close enough connected components and isolates at all levels of activity, and networks become less transitive, but more assortative for larger subgraphs. We also analyze the association between user activities and characteristics that may influence user behavior to spread information during a crisis. Users who are centrally placed in the network, less eccentric and have higher degrees, they are more active in spreading information. Our analyses provide insights on how to exploit user characteristics and network properties to spread targeted information in major disasters.

1. Introduction

Communities all over the world are frequently facing disasters in many forms (Nelson, 2013). Natural disasters alone, over the past three decades, have caused billions of dollars in property damage and killed 2.5 million people (Emergency Events Database, n.d.; Klomp and Valckx, 2014). The National Academies Committee on Increasing National Resilience to Hazards and Disaster established hazards resilience as a national imperative at all levels (personal, local, state, and national). Disaster resilience has received more emphasis in the domains of physical infrastructure systems and operations (Cutter et al., 2013); however, resilience should also incorporate social dimensions (Sadri et al., 2018). To promote disaster resilience and minimize the adverse impacts, communities need to have sufficient preparation and information to respond to an upcoming crisis (Diffenbaugh and Field, 2013; Guikema, 2009; Kennedy, 2002; Press and Hamilton, 1999). Early detection of influential agents in a crisis communication network can contribute towards relaying targeted, relevant, and timely information to the vulnerable communities and concerned population-groups. As a result, any possible disruption of information flow in the network can be resisted in such a crisis ahead of time.

Effective information dissemination constitutes the key to spread awareness to every individual in a community (Cutter and Finch, 2008; Helbing, 2013; Vespignani, 2009). It requires systematic planning, collection, organization, and delivery techniques before circulating to the target audience using different media and communication means. Online social media (such as Facebook, Twitter etc.), unlike traditional ones, can serve as alternative platforms to disseminate information during disasters. Studies have acknowledged the potential and need to efficiently analyze, record and utilize the large-scale and rich information available from these online information sources (Lazer et al., 2009). Examples of such applications can be found in many empirical studies related to emergency response (Bagrow et al., 2011; Guan et al., 2016; Hughes and Palen, 2009; Li and Rao, 2010; Van Hentenryck, 2013; Wang et al., 2014; Watts et al., 2013), crisis informatics (Caragea et al., 2011; Earle et al., 2012; Freeman, 2011; Guy et al., 2010; Pickard et al., 2011; Sakaki et al., 2010; Skinner, 2013; Ukkusuri et al., 2014), and many others (Li et al., 2016; Liang et al., 2015; Shin et al., 2016; Weng et al., 2013; Wood et al., 2013; Zhong et al., 2016). Moreover, social media connectivity and activity allow researchers to analyze and predict what happens in the real world via social network amplification (Korolov et al., 2015; Kryvasheyeu et al., 2016).

* Corresponding author.
  E-mail addresses: asadri@fiu.edu, (A.M. Sadri), samiul.hasan@ucf.edu, (S. Hasan), sukkusur@purdue.edu, (S.V. Ukkusuri), cebrian@mpib-berlin.mpg.de. (M. Cebrian).

Social science studies have reported that psychological and social factors are very important in translating hazard warning information into a collective decision (Gladwin et al., 2007; Lindell et al., 2005; Lindell and Prater, 2007a, 2007b). Evacuation studies have found significant correlations of local authorities, peers, local and national media, and internet with evacuation (Lindell et al., 2005). During Hurricane Sandy, for instance, social media played an important role on information sharing. Residents from New York and New Jersey were able to receive information on smartphones using social media as they had limited access to traditional sources of information (radio, television and others) (Kaufman et al., 2012). In areas without power, communications via online social media continued during and after the storm based on the continuous distribution of tweets observed throughout the city. Individuals were more likely to evacuate if they relied on social media for weather-related information during Sandy (Sadri et al., 2017). Although social media data has been analyzed for many disaster studies, a key question remains open: what is the role of the underlying network structure in spreading information in social media during disasters?

The interdependence between network topology and the function of network agents has important consequences on the robustness and resilience of real networks as they respond to random failure, targeted attacks or any other external perturbations (Albert et al., 2000). Understanding the coupled dynamics between network structure and function has manifold applications in various fields including infrastructure systems, supply chain and logistics, biology, social and financial systems, information and communication networks, and many others (Boccaletti et al., 2006; Newman, 2003; Ukkusuri et al., 2016). This joint association of network structure with the entities also allows the experiment of highly dynamic behavior of the network agents that exist and interact within the complex architecture. Complex network approaches have been used in many empirical studies of real world systems, such as, disease transmission (Anderson et al., 1992; Murray, 2002); transmission of computer viruses (Balthrop et al., 2004; Newman et al., 2002); collapse in financial systems (Sornette, 2009), failures of power grid (Kinney et al., 2005; Sachtjen et al., 2000); information diffusion through social networks (Coleman et al., 1966), and many others.

In this study, we investigate information spreading activities and associated network properties in social media during disasters. We have analyzed active Twitter subgraphs to understand information spreading activities of Twitter users during Hurricane Sandy. Multiple Twitter subgraphs have been created based on user activities and followee list during Hurricane Sandy. We analyze different structural properties of the subgraphs following the concepts of network science. This study contributes towards a better

understanding of user activities and interactions in social media platforms during a major crisis such as Hurricane Sandy. The information spreading patterns will be useful for the early detection of influential network agents in such crisis. Our tobit regression indicates nodes being less eccentric, more central, and having larger degrees are more capable of spreading relevant information (Table 1). Such nodes, because of higher reachability to many other nodes, can make meaningful contributions in crisis contagion and help disseminate early awareness in the hurricane prone regions.

2. Review of network science

Many new network concepts, properties and measures have been developed by running experiments on large-scale real networks. A number of statistical properties and unifying principles of real networks have been identified from these studies. Significant amount of research efforts have helped to develop new network modeling tools, reproduce the structural properties observed from empirical network data, and design such networks efficiently with a view to obtaining more advanced knowledge of the evolutionary mechanisms of network growth (Hasan and Ukkusuri, 2011). Many real networks possess interesting properties unlike random graphs indicative of possible mechanisms guiding network formation and ways to exploit network structure with specific objectives (Newman, 2003). Some of these properties, common across many real networks, are described below:

Small-world property: This property refers to the existence of relatively short paths between any pair of nodes in most networks despite their large size. The existence of this property is evident in many real networks (Milgram, 1967; Travers and Milgram, 1969; Watts and Strogatz, 1998). The small-world effect has important implications in explaining dynamics of processes occurring on real networks. In case of spreading information or ideas through a network, the small-world property suggests that the propagation will be faster on most real world networks because of short average path lengths (Newman, 2003). Three important measures to explain this property are eccentricity, radius and diameter. While the eccentricity of a node in a graph is the maximum distance (number of steps or hops) from that node to all other nodes; radius and diameter are the minimum and maximum eccentricity observed among all nodes, respectively.

Degree distributions: The degree of a node (k) is the number of direct links to other nodes in a graph. The degree distribution $P(k)$ in real networks (probability that a randomly chosen node has degree k, issignificantly different from the Poisson distribution, typically assumed in the modeling of random graphs. In fact, real networks exhibit a power law (or scale-free) degree distribution characterized by higher densities of triangles (cliques in a

Table 1
Tobit regression for activity frequency.

	N	Mean	S.D.	Min.	Max.
Activity frequency	152,933	24.87351	37.59742	10	2267
Degree	152,933	148.7643	484.8861	1	32,406
In-degree	152,933	94.80183	219.0143	0	9640
Out-degree	152,933	94.80183	447.848	0	32,397
Clustering coefficient	152,933	0.195215	0.13114	0	1
Eccentricity	152,933	5.790215	0.456218	5	8
Avg. neighbor degree	152,933	2745.791	2326.095	1.5	32,406
Betweenness centrality	152,933	1.14E−05	0.000231	0	0.037856
Closeness centrality	152,933	0.369475	0.040014	0.178652	0.530577
Eigenvector centrality	152,933	0.00109	0.002313	8.86E−13	0.071367
Degree centrality	152,933	0.000973	0.003171	6.54E−06	0.211898

Activity frequency: Tobit regression

	Coeff.	Std. err.	t-Stat	[95% Conf. int.]	
Constant	18.92990	2.73118	6.93	13.57685	24.28295
In-degree	0.00917	0.00057	16.10	0.008053	0.010285
Out-degree	0.00658	0.00026	25.63	0.006081	0.007088
Eccentricity	−5.53433	0.30563	−18.11	−6.13336	−4.9353
Closeness centrality	87.82817	3.54151	24.80	80.88689	94.76944
No. of observations	152,933				

social network, for example) (Barabási and Albert, 1999). In addition, many real networks also exhibit significant correlations in terms of node degrees or attributes. This scale-free property validates the existence of hubs, or a few nodes that are highly connected to other nodes in the network. The presence of large hubs results in a degree distribution with long tail (highly right-skewed), indicating the presence of nodes with a much higher degree than most other nodes. For an undirected network, the degree distribution $P_{degree}(k)$ can be written as follows:

$$P_{degree}(k) \propto k^{-\gamma} \tag{1}$$

where γ is some exponent and $P_{degree}(k)$ decays slowly as the degree k increases, increasing the probability of obtaining a node with a very high degree. Networks with power-law distributions are called scale-free networks (Albert and Barabási, 2002) that hold the same functional form (power laws) at all scales. The power law $P_{degree}(k)$ remains unchanged (other than a multiplicative factor) when rescaling the independent variable k by satisfying:

$$P_{degree}(xk) = x^{-\gamma} P_{degree}(k) \tag{2}$$

The presence of hubs that are orders of magnitude larger in degree than most other nodes is a characteristic of power law networks. In this study, we test the scale free property both for the activity frequency of all active nodes and the degree distribution of subgraphs being active at different activity levels.

Transitivity: This property is a distinctive deviation from the properties of random graphs. Network transitivity implies that two nodes are highly likely to be connected in a network, given each of the nodes are connected to some other node. This is indicative of heightened number of triangles that exist in real networks (sets of three nodes each of which is connected to each of the others) (Newman, 2003). The existence of triangles can be quantified by *Clustering Coefficient, C*:

$$C = \frac{3 * Number\ of\ triangles\ in\ the\ network}{Number\ of\ connected\ triples\ of\ nodes} \tag{3}$$

A *connected triple* refers to a single node with links running to an unordered pair of others. In case of social networks, transitivity refers to the fact that the friend of one's friend is likely also to be the friend of that person. Another important notion is *Network Density*, frequently used in the sociological literature (Scott, 2012). The density is 0 for a graph without any link between nodes and 1 for a completely connected graph.

Network resilience: This property, related to degree distributions, refers to the resilience of networks as a result of removing random nodes in the network and the level of resilience to such vertex removal varies across networks depending on the network topology (Newman, 2003). Networks in which most of the nodes have low degree have less disruption since these nodes lie on few paths between others; whereas removal of high degree nodes in a large real network can result in major disruption. The usual length of these paths will increase if nodes are removed from a network, resulting in disconnected pairs of nodes and making it more difficult for network agents to communicate.

Node-level properties: The node degree is the number of edges adjacent to that node (deg_i). In-degree is the number of edges pointing in to the node (in_deg_i) and out-degree is the number of edges pointing out of the node (out_deg_i). Average neighbor degree refers average degree of the neighborhood ($z_{n, i}$) of each node i is:

$$z_{n,i} = \frac{1}{|N_i|} \sum_{j \in N_i} z_j \tag{4}$$

where, $N(i)$ are the neighbors of node i; z_j is the degree of node j that belongs to N_i. In case of weighted graphs, weighted degree of each node can

be used (Barrat et al., 2004). In case of an unweighted graph, the clustering coefficient (cc_i) of a node i refers to the fraction of possible triangles that exist through that node:

$$cc_i = \frac{2\ T_i}{deg_i * [\ deg_i - 1]} \tag{5}$$

where, T_i is the number of triangles that exist through node i and deg_i is the degree of node i. In case of weighted graphs, this clustering coefficient can be defined as the geometric average of the sub-graph edge weights (Saramäki et al., 2007). The eccentricity of node i is the maximum distance from node i to every other nodes in the graph G (ecc_i).

Out of a number centrality measures, betweenness centrality (BC_i) of node i is the sum of the fraction of all-pairs of shortest path that pass through node i:

$$BC_i = \sum_{x,y \in V} \frac{\theta_{(x,y\ |\ j)}}{\theta_{(x,y)}} \tag{6}$$

where, V is the set of nodes in G, $\theta_{(x,y)}$ is the number of shortest (x,y) paths, and $\theta_{(x,y\ |\ j)}$ is the number of paths that pass through some node j other than (x,y). Please refer to Brandes (2001, 2008) and Brandes and Pich (2007) for more details. The closeness centrality (CC_i) of node i is the reciprocal of the sum of the shortest path distances from node i to all $(n - 1)$ other nodes in the graph G:

$$CC_i = \frac{n-1}{\sum_{j=1}^{n-1} \theta_{(j,i)}} \tag{7}$$

where, $\theta_{(j,i)}$ is the shortest path distance between node j and node i and n is the number of total nodes in graph G. Closeness is normalized by the sum of minimum possible distances of $(n - 1)$ since the sum of the distances depends on the number of nodes in the graph. Higher values of closeness imply higher centrality. Please refer to Freeman (1978) for details. The eigenvector centrality (EC_i) computes the centrality for a node i based on the centrality of its neighbors. The eigenvector centrality for node i is:

$$A\,x = \lambda\,x \tag{8}$$

where A is the adjacency matrix of the graph G with eigenvalue λ. Perron–Frobenius theorem suggests that there is a unique and positive solution if λ is the largest eigenvalue associated with the eigenvector of the adjacency matrix A (Bonacich, 1987; Newman, 2010). Finally, degree centrality for a node is just the fraction of nodes it is connected to.

Other network properties: Some other common properties are observed in many real networks such as mixing patterns (selective linking), network homophily or similarity, degree correlations, preferential attachment, community structure, network navigation, size of giant components among others (Newman, 2003).

3. Data description

In 2012, residents in the coastal areas of New York and New Jersey experienced a massive storm surge produced by Hurricane Sandy, a late season hurricane causing about $50 B in property damage, 72 fatalities in the mid-Atlantic and northeastern United States, and at least 147 direct deaths across the Atlantic basin (Blake et al., 2013). Sandy's wind and flood are the key contributors of the heightened number of fatalities (Gladwin et al., 2013). In addition, 570K buildings were destroyed, 20K flights were cancelled, and 8.6M power outages in 17 states among other direct impacts of Sandy (Halverson and Rabenhorst, 2013). Moreover, thousands of people were displaced from their homes (Abramson and Redlener, 2012) and 230K cars were destroyed by the floods even though the residents were given early warnings about the oncoming storm and the likely impact (Meyer et al., 2014). The specific date, time, location

Table 2
Specific events along the path of Hurricane Sandy.

Date	Time	Nearby location	Event
October 22, 2012	12:00 UTC	Kingston, Jamaica	Sandy formed and officially assigned name
October 24, 2012	19:00 UTC	Jamaica	First landfall as a Category 1 hurricane
October 25, 2012	05:30 UTC	Cuba	Second landfall as a Category 3 hurricane
October 29, 2012	12:00 UTC	Atlantic City	Re-intensified to the maximum wind speeds
October 29, 2012	23:30 UTC	Near Brigantine in New Jersey	Final landfall as a post-tropical storm

and event, attributed to Hurricane Sandy, are presented in Table 2. Please see Kryvasheyeu et al. (2015) for further details.

Twitter users can share short messages up to 140 characters and follow other users creating a network of a large number of accounts with characteristics both of a social network and an informational network (Myers et al., 2014). As of 2017, messages can contain up to 280 characters, as such the Hurricane Sandy tweets analyzed in this study was significantly shorter than current tweets. The social network properties of Twitter provide access to geographically and personally relevant information and the information network properties instigate information contagion globally (Kryvasheyeu et al., 2016). These specific features make Twitter particularly useful for effective information dissemination during crises. From an emergency research perspective, many researchers used Twitter to study the service characteristics (Guy et al., 2010; Li and Rao, 2010), retweeting activity (Kogan et al., 2015; Starbird and Palen, 2010), situational awareness (Power et al., 2014; Vieweg et al., 2010), online communication of emergency responders (Hughes et al., 2014; St Denis et al., 2014), text classification and event detection (Caragea et al., 2011; Earle et al., 2012; Imran et al., 2013; Kumar et al., 2014; Sakaki et al., 2010), devise sensor techniques for early awareness (Kryvasheyeu et al., 2015), quantifying human mobility (Wang and Taylor, 2014, 2015), and disaster relief efforts (Gao et al., 2011).

In this study, we analyze raw data (~52 M tweets, ~13 M users, Oct 14–Nov 12, 2012) obtained from Twitter. Please see Kryvasheyeu et al. (2015) for the detailed steps involved in data collection. The data includes a text database with user and text identifiers, texts, and some additional useful information. The network database includes the relationship graphs of active users i.e. the list of followees for each user. These were reconstructed using Twitter API. Only a minor fraction of the texts (~1.35%) are geo-tagged by Twitter. For relevance, user activity was assessed on the basis of number of tweets (~11.83 M) in the data that included the word 'sandy', co-appeared with other words after filtering out ~46.45 M tweets that are in English i.e. non-English tweets were removed. Here, the rationale is certain keywords, such as "storm", "power", "gas", etc. are contaminated by irrelevant messages simply due to their convoluted nature. For example, the keyword "storm" may belong to small-scale local weather events. Similarly, "power" can be used in the context of politics and "gas" may refer to regular user concerns about gas. Strict relevance filtering to avoid "false positives" (messages with no relevance to Hurricane Sandy) also helps to suppress noise messages prior to the formation of the hurricane (Kryvasheyeu et al., 2015).

Thus, relevance filtering was carefully done before applying network models on massive subgraphs of Twitter users being concerned about Hurricane Sandy. For the analysis of directed graphs, we considered a network link received by user from his followee. From Twitter perspective, a followee is the user who is being followed by another user and the information flows from the followee to the followers. In the network data, we observed that a number of highly active nodes did not have any followee, however they appeared in the followee list less active users which is indicative of the direction and rate of information flow. Some active users did not appear in the network database for which we assumed zero followee since the current length of followee list on Twitter is close to zero, even after three years of data collection.

4. Methods and results

4.1. Subgraph construction

Given the scope, this study is focused on exploring the social media user interaction network properties in a major disaster. As such, Twitter data was obtained covering the warning, response and early recovery phases of Hurricane Sandy. Since it is almost impossible to observe the complete graph of Twitter especially and we focused only on the Sandy related communication patterns, our network analyses were based on the subgraphs of Twitter that included only users who interacted about Hurricane Sandy. In order to create subgraphs at any given activity level (Bondy and Murty, 1976), we first observe the followees of an active user and identify all the active followees of that user. We construct a directed subgraph of all active users having links from user followees directed towards the active nodes. We then observe the association between the frequencies of user activity (i.e. number of tweets during the analysis period) and network properties (both global and local) by running networks models for these subgraphs. Global i.e. graph-level network properties that we assessed include number of nodes and links (both for the graph and the largest connected component); number of isolates and connected component; mean degree, radius and diameter; network density, transitivity, and assortativity. Local i.e. node-level properties include node activity, degree, average neighbor degree, clustering coefficient, eccentricity, and different measures of centrality. The larger the size of the subgraphs, the more nodes it includes from a lower activity level. Fig. 1 visualizes the subgraphs of different sizes and their largest connected components. Network visualization shows highly active nodes in the active subgraphs appearing both at the largest connected component and as isolates in the periphery (Fig. 1a). Within the largest connected component, highly active nodes appear at different positions (Fig. 1b-d). The ego node of the largest hub depicts its influential position in the subgraph connectivity directing our attention towards a node-level analysis of the subgraphs (Fig. 1e).

We followed the following definition (Bondy and Murty, 1976) to construct the subgraphs:

A graph G is an ordered triple $(V(G), E(G), \psi_G)$ consisting of a non-empty set $V(G)$ of nodes, a set $E(G)$ of links being disjoint from $V(G)$ and an incidence function ψ_G that associates with each edge of G an unordered pair of (not necessarily distinct) vertices of G. If e is an edge and u and v are vertices such that $\psi_G(e) = uv$, then e is said to join u and v; the nodes u and v are called the ends of e. A graph H is a subgraph of G $(H \subseteq G)$ if $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$, and ψ_H is the restriction of ψ_G to $E(H)$. When $H \subseteq G$ but $H \neq G$, we write $H \subset G$ and call H a proper subgraph of G. If H is a subgraph of G, G is a supergraph of H.

4.2. Activity and degree distributions

We analyze the relationship between user activity and the corresponding degree distributions. We obtain the best fitting to the user activity and subgraph degree distributions (see Fig. 2 for details) and a value of x_{min} which refers to the minimal value of x at which the power law begins to become valid (Alstott et al., 2014). User activity frequency based on all relevant keywords follows a power law distribution ($\gamma = 2.71 \pm 0.005$; $p < 0.001$; $x_{min} = 39$), whereas, activity frequency (AF) based on keywords co-appeared with 'sandy' follows a truncated power law distribution ($\gamma = 2.795 \pm 0.016$; $p < 0.001$). This indicates the existence of few nodes
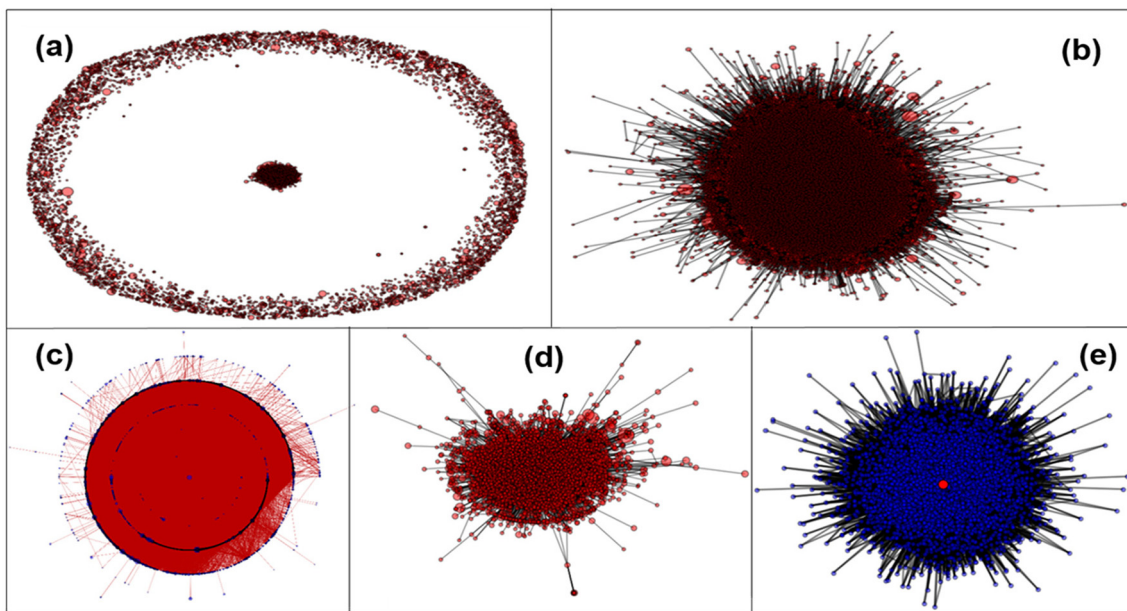
Fig. 1. Snapshots of active subgraphs with their elements. (a) Full subgraph (~0.16 M nodes, ~14.50 M links, AF ≥ 10, ~3.92 M tweets); (b) largest connected component of the subgraph (~0.15 M nodes, ~11.38 M links, AF ≥ 10); (c) circular tree visualization of largest connected component (~12 K nodes, ~0.50 M links, AF ≥ 50); (d) regular visualization of largest connected component (~12 K nodes, ~0.50 M links, AF ≥ 50); (e) largest hub (AF ≥ 50). ***Node size is proportional to node activity in each case.

capable of spreading information quickly while many other nodes being less active. The degree distributions of the subgraphs at different activity frequency (AF) levels follow a truncated power law ($\gamma = 3.057 \pm 0.067$; $p < 0.01$; $AF \geq 10$). Here, γ is the slope of the distribution. This replicates the scale-free property of many real networks having fewer nodes with larger degrees and many nodes having relatively low degree. When γ is high, the number of nodes with high degree is smaller than the number of nodes with low degree. We may thus think that a low value of γ denotes a more equal distribution, and higher values of γ denote more and more unfair degree distributions.

However, this might not be the case and the opposite may become true i.e. a high value of γ represents a network in which the distribution of edges is fairer. The best fit power law may only cover a portion of the distribution's tail (Alstott et al., 2014). There are domains in which the power law distribution is a superior fit to the lognormal (Klaus et al., 2011). However, difficulties in distinguishing the power law from the lognormal are common and well-described, and similar issues apply to the stretched exponential and other heavy-tailed distributions (Malevergne et al., 2009; Malevergne et al., 2005). Our comments on the distributions fitting are based on pairwise comparison between power law, truncated power law, lognormal, and exponential distributions.

4.3. Network analyses

Fig. 3 shows the variation of the subgraph network properties at various activity levels. It is important to note here that the larger the size of the subgraphs, the more nodes it include from a lower activity level. We observe that the number of nodes and links generated in these subgraphs (both directed and undirected) grow exponentially for larger subgraphs. A similar pattern is observed for the nodes and links that exist in the largest connected component. There exists almost equal number of connected components and isolates for all levels of activity. Network densities of the subgraphs (both directed and undirected) tend to zero for larger subgraphs, having slightly higher densities in the largest connected component in each case. This implies that the connectivity between nodes do not follow the rate at which the network grows for larger subgraphs.

Network transitivity implies the probability of any two given nodes in the graph to be connected if they are already connected to some other node. The average clustering coefficient of the undirected subgraphs range from 0.2 to 0.4 and decreases with the size of the subgraphs (see Fig. 4). The network transitivity, based on average clustering coefficient, suggests that the subgraphs become less transitive as their size grows. The increase in average degree of the nodes is indicative of more nodes that are reachable in larger subgraphs on average. The degree Pearson correlation coefficient approaches 0 for larger subgraphs. This is a measure of graph assortativity in terms of node degree and a network is said to be assortative when high degree nodes are, on average, connected to other nodes with high degree and low degree nodes are, on average, connected to other nodes with low degree. Since degree assortativity measures the similarity of connections in the graph with respect to the node degree, we observe that the networks become more assortative for larger subgraphs. While the eccentricity of a node in a graph is the maximum distance (number of steps or hops) from that node to all other nodes; radius and diameter are the minimum and maximum eccentricity observed among all nodes, respectively. For a larger subgraph, we observe that the radius takes a constant value of 5, while diameter approaches 8.

Node level properties are important to understand the role and contribution of different nodes (network agents) on the information propagation at a local scale. To obtain node level properties, we first construct an active subgraph with activity frequency (AF) ≥ 10 that includes a directed graph of 157,622 nodes and 14,498,349 links. Then we run different network models to obtain node-level properties of the undirected largest connected component with 152,933 nodes and 11,375,485 links. We observe that most of these nodes had a degree close to ~25 with activity frequency around 13. While some of the nodes, having equivalent degree, were highly active; most of the nodes in this degree region remained less active. We observe fewer nodes in the higher degree zones who remain less active than some lower degree nodes (Fig. 5a–d). However, these nodes can play important role during a crisis or emergency because of their higher access to many other nodes. Similar but less smooth trend was observed with respect to average neighbor degree. This node property is related to assortativity that measures the similarity of connections in the graph with respect to the node degree. An important insight here is that we see a chunk of nodes having very high degree neighbors who remained less active (Fig. 5a–b). Since high degree nodes are more likely to be active individually, nodes having high degree neighbors would be less summoned by
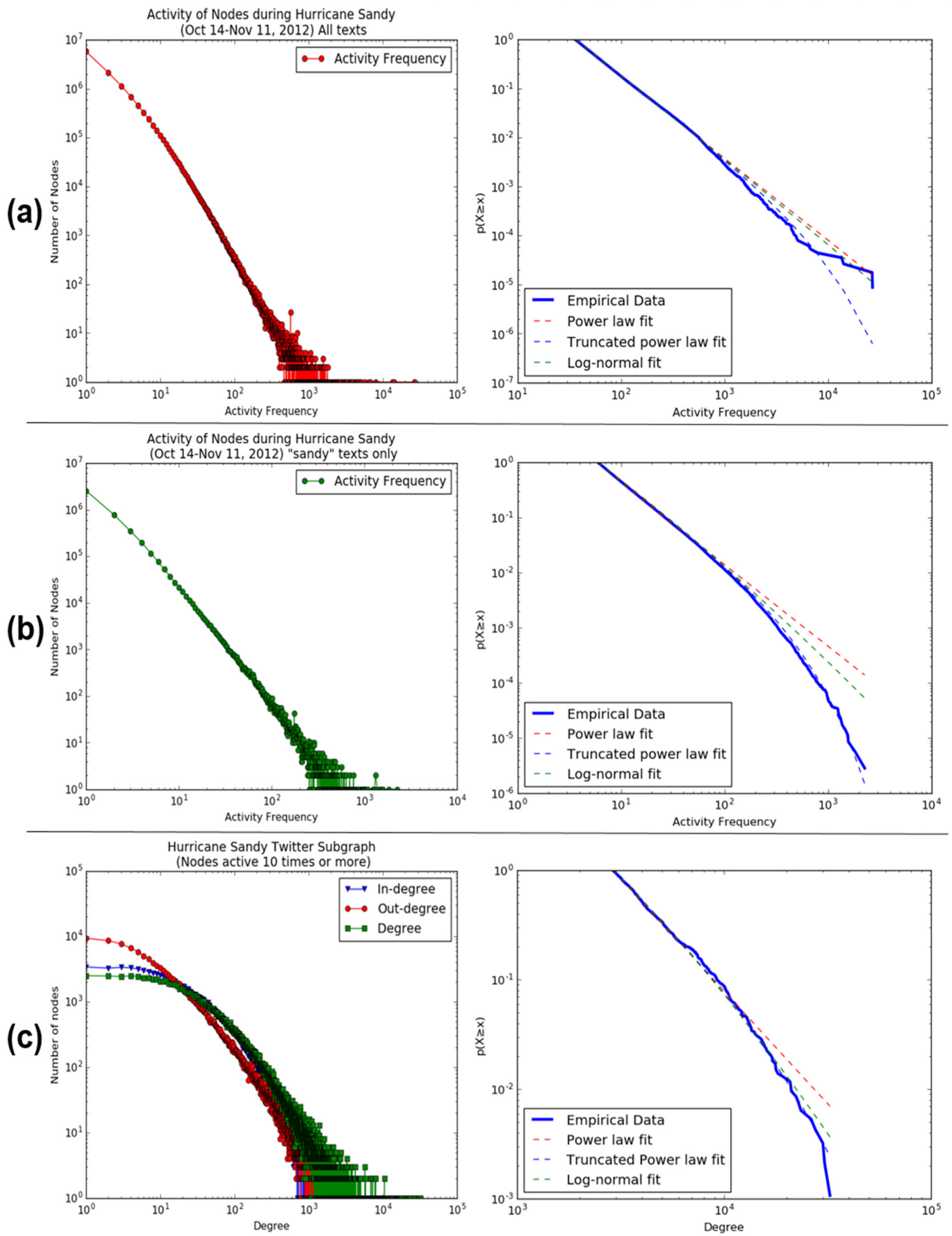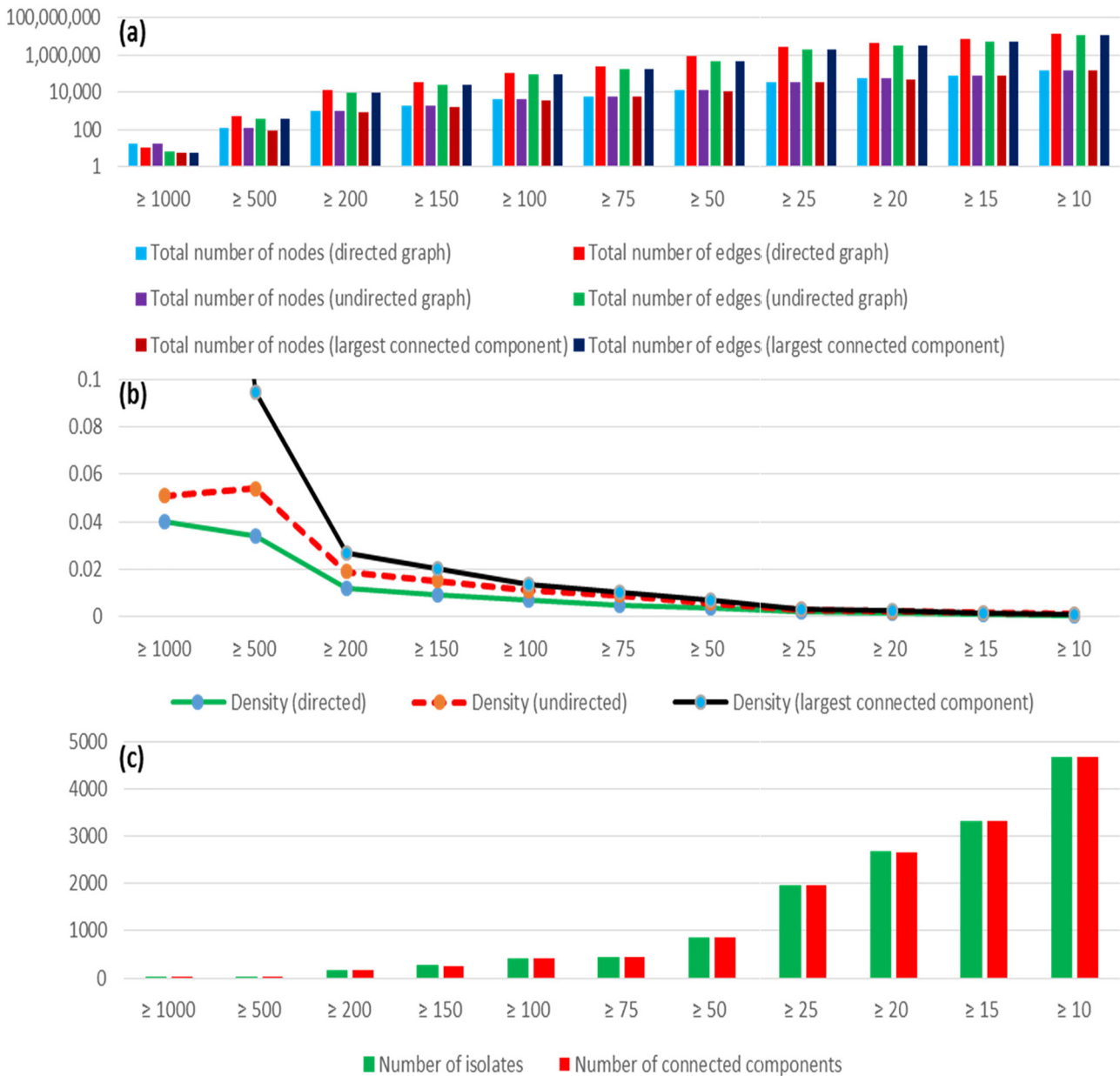
(caption on next page)

Fig. 3. Subgraph properties at different activity levels (where activity level is defined by the number of tweets made by a node). (a) Number of nodes and links, (b) network densities, (c) number of isolates and connected components.

their neighbors because those neighboring nodes already self-sufficient in disseminating or obtaining information. On the other hand, those nodes having degree neighbors might need extra time to process oversaturated information from highly active neighbors.

The association of activity frequency with node-level clustering co-efficient and eccentricity (Fig. 6a–b) also show well-defined range. Since eccentricity of a node is the maximum distance from that node to all other nodes, we observe that most of the nodes had an eccentricity of 6, many of them remained less active while only a few of them were highly active. More importantly, for some nodes having less eccentric-ity, we observed their rigidity to be less active during crisis. Fig. 6a–b shows that many nodes, even being part of the largest connected

component, did form any cluster and remained less active. These nodes are less reachable from the nodes who are more central and form clusters. Such nodes should be given due consideration for effec-tive information dissemination during crises. Turning to the centrality measures (Fig. 6c–d), we observe that only closeness centrality shows a well-defined range (Fig. 6c–d). Degree centrality and eigenvector cen-trality shows similar patterns. Betweenness centrality suggests that al-most all the nodes were having centralities equal to zero in terms of their betweenness in the network. The key take away from the centrality parameters is the pattern presented by the closeness centrality which is indicative of a lot of nodes being highly central in terms of their close-ness with many other nodes who remained significantly less active.

← **Fig. 2.** Activity and degree distributions. (a) Activity distribution based on all tweets after initial filtering (~46.45 M tweets). This follows a power law distribution. (b) Activity distribution of different users after the 'sandy' filtering (~11.83 M tweets). This follows a truncated power law distribution. (c) Degree distribution of the largest directed subgraph (~0.16 M nodes, ~14.50 M links, AF ≥ 10, ~3.92 M tweets). This also follows a truncated power law distribution.
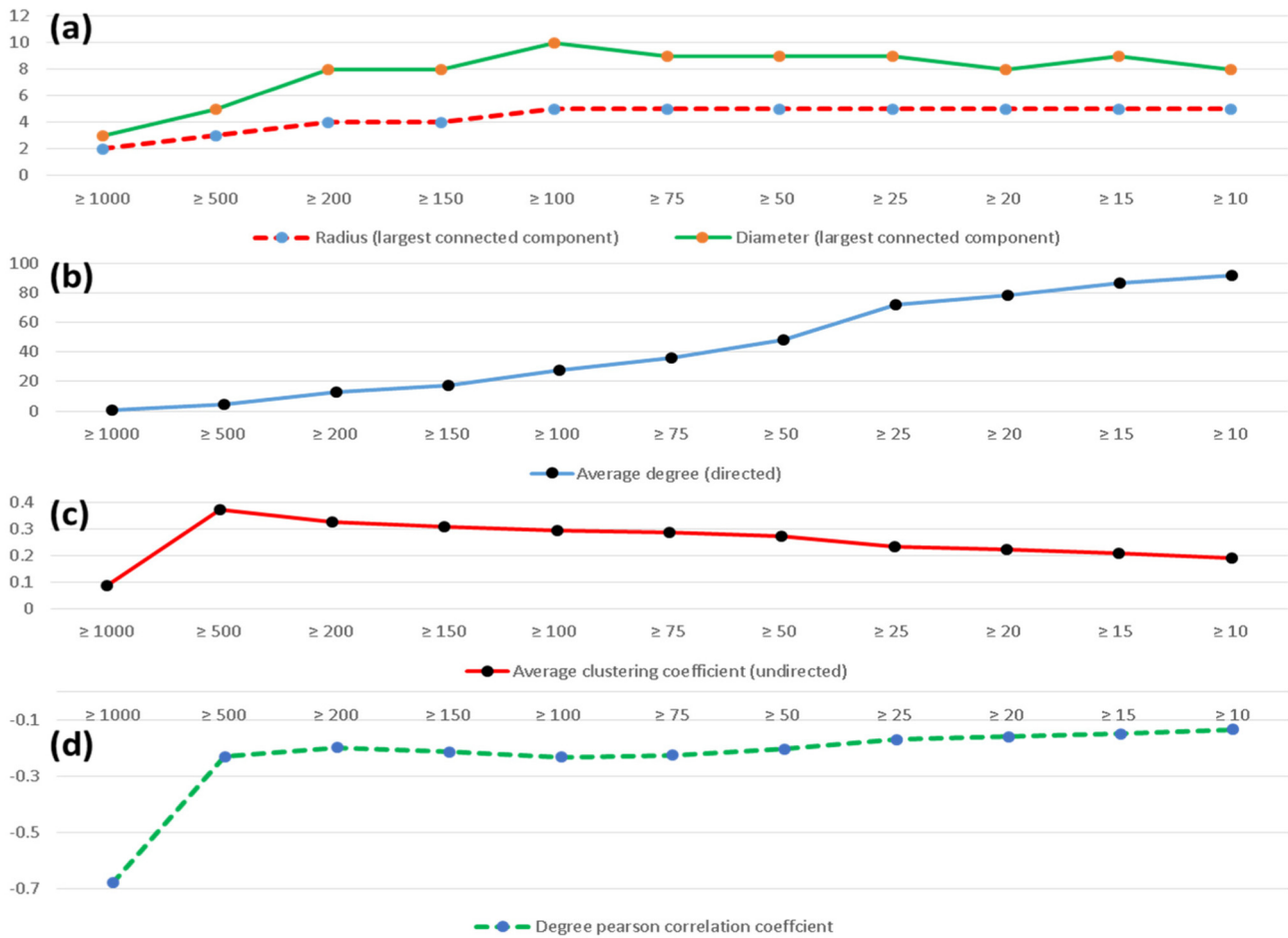
Fig. 4. Subgraph properties at different activity levels. (a) Radius and diameter of the largest connected component, (b) average degree, (c) average clustering coefficient, and (d) the degree Pearson correlation coefficient.

4.4. Information spreading activity of network agents

To assess how network agents performed in terms of spreading relevant information about Sandy, we fit both linear and quadratic models of the form:

$$y \sim \theta_0 + \theta_1\ x + \epsilon \tag{9}$$

$$y \sim \theta_0 + \theta_1\ x + \theta_1\ x^2 + \epsilon \tag{10}$$

respectively, where y represents activity frequency of nodes and x is a node-level network attribute. This has been done by observing node-level network properties of the largest directed subgraph AF ≥ 10 that includes 157,622 active nodes originally and 152,933 active nodes in the largest connected component. We examined the effects of degree, in-degree, out-degree, eccentricity, and closeness centrality on spreading capacity i.e. frequency of relevant tweeting activity (see Fig. 7). The quadratic model fit the data well in each case and was chosen over the linear model. Fig. 7a–e shows the maximum likelihood fit of both the quadratic and linear model (dashed line), where the shaded area is the 95% confidence interval of the quadratic model, and the dots show the performance of each of the network agents. In addition to this univariate analysis, we also run a multivariate tobit regression to determine the combined effects of network variables on information spreading. Since activity frequency (AF) is left-censored at 10, we used tobit regression which is designed for left-censored dependent variables. We report (Table 1) the mean, standard deviation, minimum, and maximum of the variables tested in the tobit regression. We observe higher variability of closeness centrality; however, the variability of other centrality measures is insignificant while their means close to being zero.

Our analysis indicates higher spreading activity for nodes having larger in-degrees and out-degrees in a directed network (larger degrees for undirected graph). This implies that the more information (links) a given node receives from other nodes, the more active is that node during crisis. On the other hand, a node is also highly likely to be influential in case of having more out-degrees i.e. links directing to other nodes that allow them to disseminate crisis information. The coefficient estimated for closeness centrality suggests more influential capability of a node by being more central in the active subgraph. Such nodes occupy a very convenient position in a network to be able to contribute highly in the information spreading dynamics. We also observed that less eccentric nodes are more capable of spreading information because of their higher reachability to any given node in the network. All the network variables tested under tobit regression are significant at $p < 0.001$.

5. Conclusions

The primary focus of this study is to understand the interdependence between network topology and activities of network agents during disasters. Social communication networks play a critical role during emergencies since people may obtain weather information from traditional media such as radio or television and social media such as Facebook, Twitter, or the internet. In this study, Twitter subgraphs have been analyzed based on user activity during Hurricane Sandy to reveal the information spreading activity of network agents and the associated network properties that evolved during this major disaster. For user activity at any given level, subgraphs of social networks were constructed from the user followee list obtained at the time of data collection. For relevance, user activity was assessed on
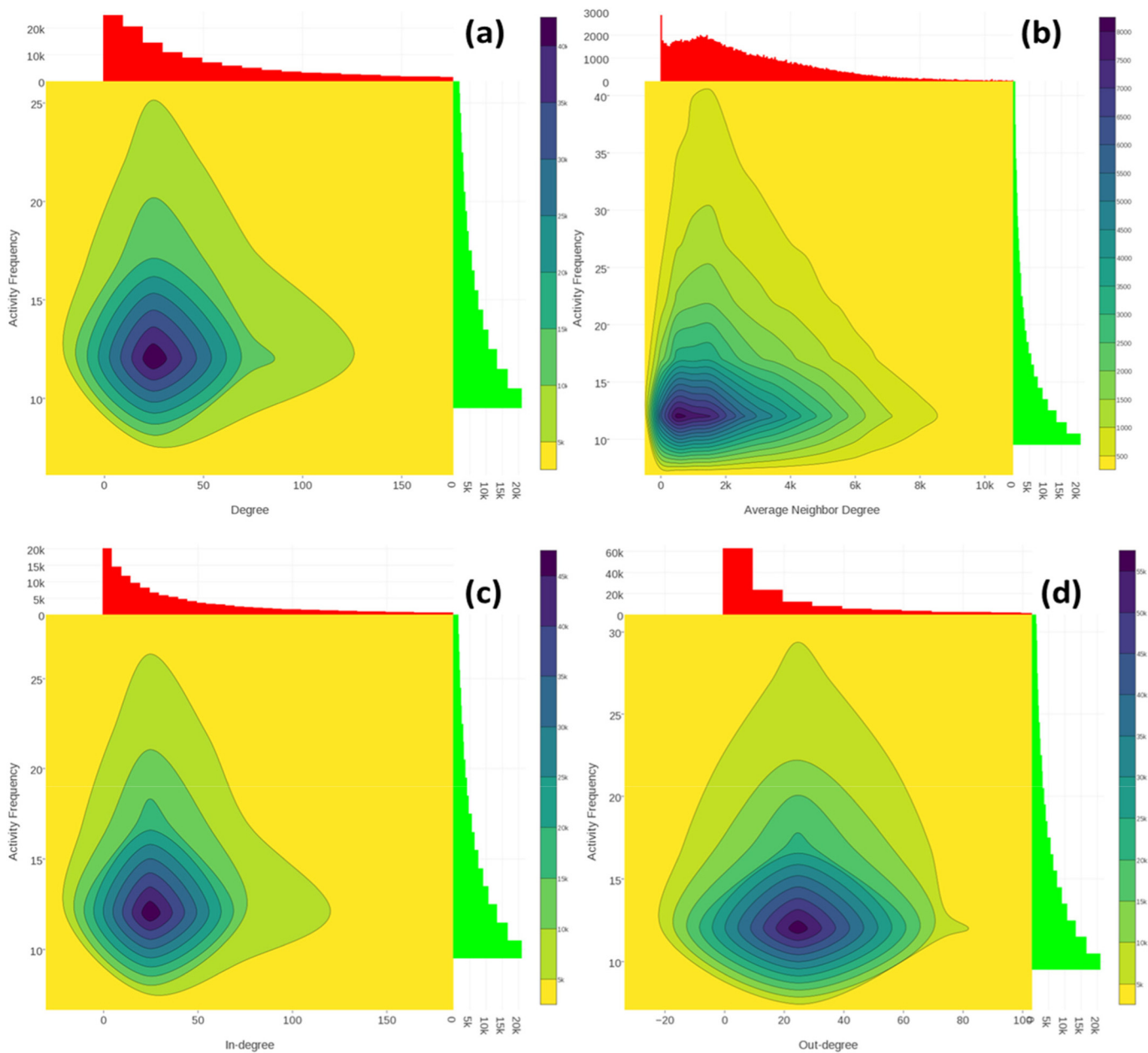
Fig. 5. Node-level properties: (a) Degree, (b) avg. neighbor degree, (c) in-degree and (d) out-degree.

the basis of number of tweets in the data that included the word 'sandy', co-appeared with other words. For the analysis of directed graphs, we considered a network link received by user from his followee. Based on our subgraph analysis at different activity levels, we reveal several information spreading characteristics of Hurricane Sandy.

We observe that information spreading activity of nodes follows a power-law showing the existence of few nodes highly active disseminating information and many other nodes being less active. The degree distributions of the communication network also follow a power-law, executing the scale-free property of many real networks (fewer nodes with larger degrees and many other nodes with fewer degrees). Network visualization shows highly active nodes in the active subgraphs appearing both at the largest connected component and as isolates in the periphery. Within the largest connected component, highly active nodes appear at different positions. The ego node of the largest hub depicts its influential position in the subgraph connectivity directing our attention towards node level analysis.

Network analysis at different activity levels suggests that the number of nodes and links in these subgraphs (both directed and undirected) grow log-linearly with the size of subgraphs that includes close enough connected components and isolates. In contrast, the overall network connectivity (i.e. subgraph densities) tends to become zero for larger subgraphs implying that having a large number of active nodes does not help much in spreading the information or awareness even though they heavily load the network. Also, the existence of significant number of network isolates at all levels does not help in crisis because of their individual activity not contributing enough to the hazard warning dissemination. For larger subgraphs, networks become less transitive, but more assortative. This implies that active network agents are more likely to connect with similar agents (for example, having similar degrees) without contributing much in forming clusters in the neighborhood at large. The radius of the largest connect components in the larger subgraphs becomes stable at 5 that is indicative of the reachability from a given node to any other node in five steps at the maximum.
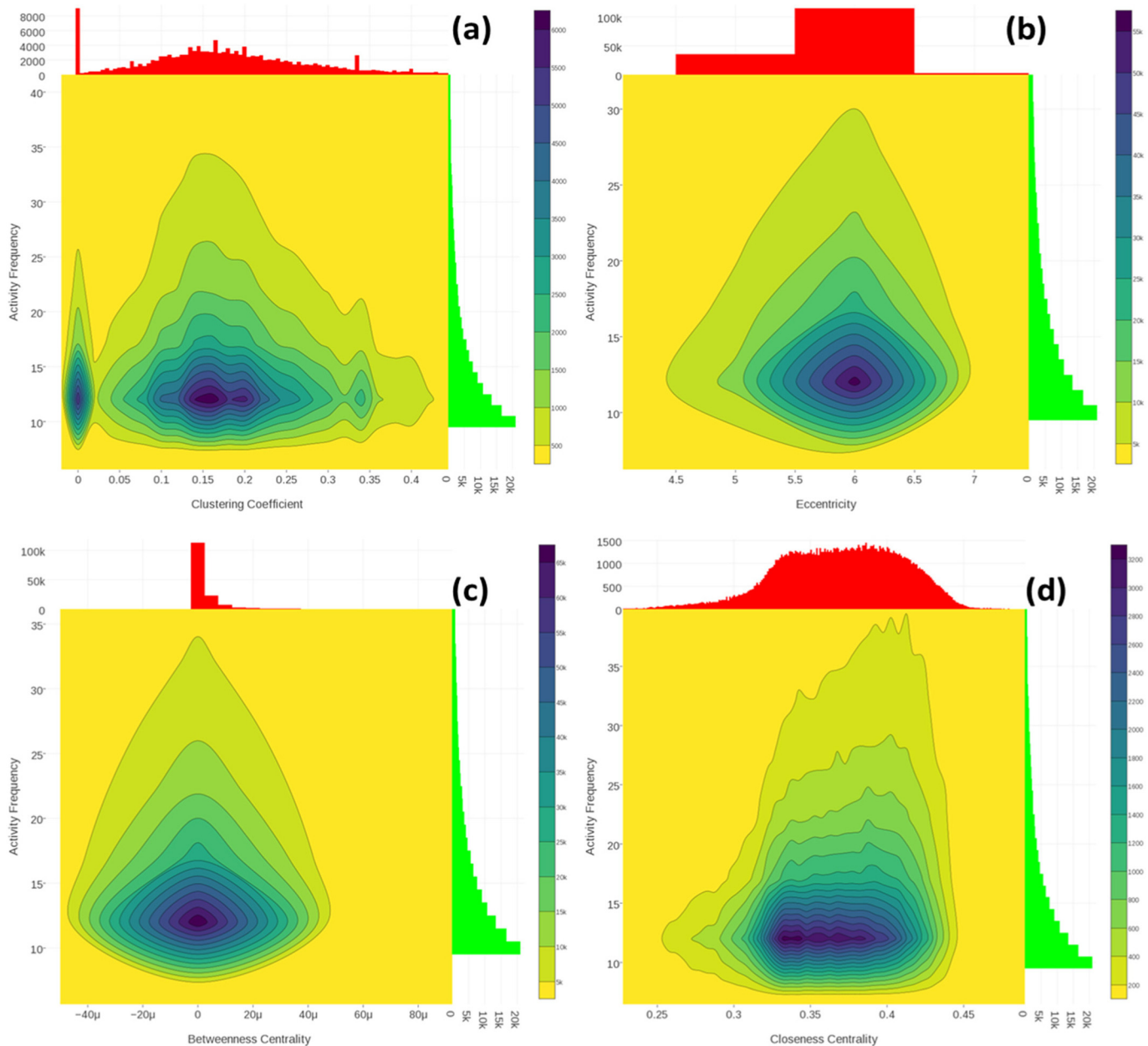
Fig. 6. Node-level properties: (a) Clustering coefficient, (b) eccentricity, (c) betweenness centrality and (d) closeness centrality.

Node-level information spreading activity of network agents was assessed by running univariate linear and quadratic models with tweeting activity as a function of several topological attributes. We examined the effects of degree, in-degree, out-degree, eccentricity, and closeness centrality on spreading capacity i.e. frequency of relevant tweeting activity. The quadratic model fit the data well in each case and was chosen over the linear model. Our analysis is indicative of higher spreading capacity for nodes having larger in-degrees and out-degrees in a directed network (larger degrees for undirected graph). A node is also highly likely to be influential in case of having more out-degrees i.e. links directing to other nodes that allow them to disseminate crisis information. Nodes are more capable of spreading information if occupying more central positions in the network and being less eccentric.

This study contributes towards a better understanding of user interactions in social media platforms during a major crisis such as Hurricane Sandy. The information spreading patterns will be useful for the early detection of influential network agents (more central, higher degrees, and less eccentric) in such crisis. From warning to evacuation to the post-storm recovery, such influential nodes can help disseminate more targeted information to reach out vulnerable communities at all phases of the disaster. For example, celebrities or political leaders typically occupy such positions on social media and may contribute to faster dissemination of relevant crisis information. The findings of this study are specific to Hurricane Sandy, future studies should validate these results with other major hurricanes and check if such insights can be generalized to other forms of disaster. The time frame in which tweets were gathered may have included a lead-up to the event as well as post-event study. Future data analysis should include breakdown of datasets (i) *by stage of event e.g. before, during, after*; (ii) *by location (geo-parsing text or using geo-tag)*; (iii) *by topic (e.g. power outage, flooding report, crises for help, damage reports)*; (iv) *posts re-posting mainstream media VS original posts among others.*
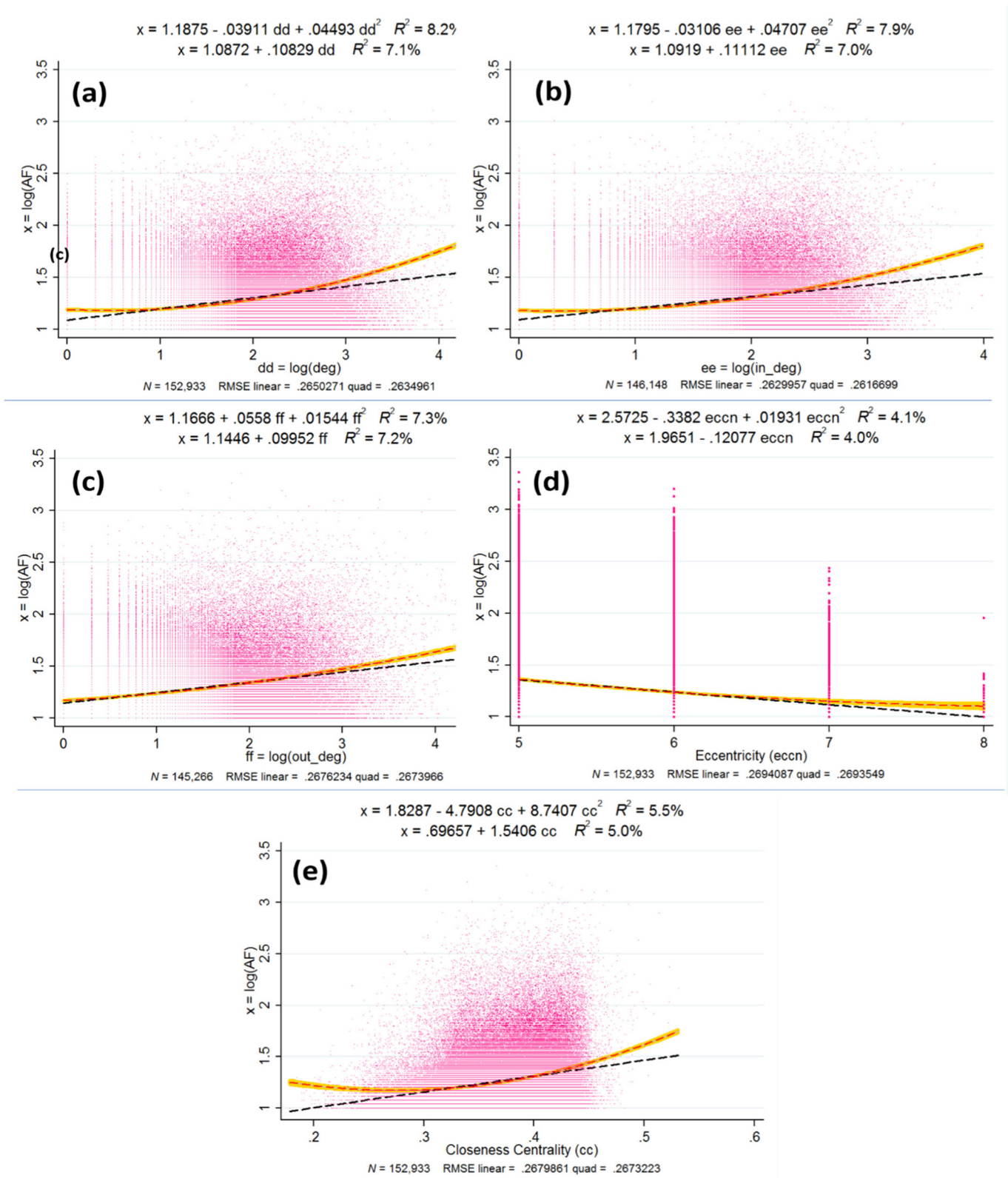
Fig. 7. Information spreading capacity. (a) Degree, (b) in-degree, (c) out-degree, (d) eccentricity and (e) closeness centrality.

Future studies should also consider the dynamics of information spreading activities of network agents. This study motivates such reproducible efforts for disaster scholars and use network science concepts to analyze disaster communication networks. Recently, transportation researchers

used these data sources extensively for problems related to human mobility patterns (Hasan et al., 2013), origin-destination demand estimation (Cebelak, 2013; Chen and Mahmassani, 2016; Jin et al., 2014; Lee et al., 2016b; Yang et al., 2014), activity-pattern modelling (Hasan and Ukkusuri,

2014, 2015; Lee et al., 2016a; Zhao and Zhang, 2016), social influence in activity patterns (Hasan et al., 2016), travel survey methods (Abbasi et al., 2015; Maghrebi et al., 2015), transit service characteristics (Collins et al., 2013), and crisis informatics (Ukkusuri et al., 2014) among others. The findings of this study also provide prominent directions towards transportation policy and practice perspectives such as identifying influential social media users as part of efficient evacuation planning and management.

CRediT authorship contribution statement

Arif Mohaimin Sadri: Conceptualization, Methodology, Formal analysis, Writing - original draft, Visualization. **Samiul Hasan:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Satish V. Ukkusuri:** Conceptualization, Writing - original draft, Supervision, Funding acquisition. **Manuel Cebrian:** Conceptualization, Resources, Data curation, Writing - original draft.

References

Abbasi, A., Rashidi, T.H., Maghrebi, M., Waller, S.T., 2015. Utilising location based social media in travel survey methods: bringing Twitter data into the play. Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-based Social Networks. ACM, p. 1.

Abramson, D.M., Redlener, I., 2012. Hurricane Sandy: lessons learned, again. Disaster medicine and public health preparedness 6, 328–329.

Albert, R, Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47.

Albert, R., Jeong, H., Barabási, A.-L., 2000. Error and attack tolerance of complex networks. Nature 406, 378–382.

Alstott, J., Bullmore, E., Plenz, D., 2014. powerlaw: a Python package for analysis of heavy-tailed distributions. PLoS One 9, e85777.

Anderson, R.M., May, R.M., Anderson, B., 1992. Infectious Diseases of Humans: Dynamics and Control. Wiley Online Library.

Bagrow, J.P., Wang, D., Barabasi, A.-L., 2011. Collective response of human populations to large-scale emergencies. PLoS One 6, e17680.

Balthrop, J., Forrest, S., Newman, M.E., Williamson, M.M., 2004. Technological networks and the spread of computer viruses. Science 304, 527–529.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. U. S. A. 101, 3747–3752.

Blake, E.S., Kimberlain, T.B., Berg, R.J., Cangialosi, J., Beven II, J.L., 2013. Tropical cyclone report: Hurricane Sandy. National Hurricane Center 12, 1–10.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., 2006. Complex networks: structure and dynamics. Phys. Rep. 424, 175–308.

Bonacich, P., 1987. Power and centrality: a family of measures. Am. J. Sociol. 1170–1182.

Bondy, J.A., Murty, U.S.R., 1976. Graph Theory With Applications. Citeseer.

Brandes, U., 2001. A faster algorithm for betweenness centrality*. J. Math. Sociol. 25, 163–177.

Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. Soc. Networks 30, 136–145.

Brandes, U., Pich, C., 2007. Centrality estimation in large networks. International Journal of Bifurcation and Chaos 17, 2303–2318.

Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H.-W., Mitra, P., Wu, D., Tapia, A.H., Giles, L., Jansen, B.J., 2011. Classifying text messages for the Haiti earthquake. Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management (ISCRAM2011) Citeseer.

Cebelak, M.K., 2013. Location-based Social Networking Data: Doubly-constrained Gravity Model Origin-destination Estimation of the Urban Travel Demand for Austin, TX.

Chen, Y., Mahmassani, H.S., 2016. Exploring Activity and Destination Choice Behavior in Two Metropolitan Areas Using Social Networking Data, Transportation Research Board 95th Annual Meeting.

Coleman, J.S., Katz, E., Menzel, H., 1966. Medical Innovation: A Diffusion Study. Bobbs-Merrill Co.

Collins, C., Hasan, S., Ukkusuri, S.V., 2013. A novel transit rider satisfaction metric: rider sentiments measured from online social media data. J. Public Transp. 16 (2).

Cutter, S.L., Finch, C., 2008. Temporal and spatial changes in social vulnerability to natural hazards. Proc. Natl. Acad. Sci. 105, 2301–2306.

Cutter, S.L., Ahearn, J.A., Amadei, B., Crawford, P., Eide, E.A., Galloway, G.E., Goodchild, M. F., Kunreuther, H.C., Li-Vollmer, M., Schoch-Spana, M., 2013. Disaster resilience: a national imperative. Environ. Sci. Policy Sustain. Dev. 55, 25–29.

Diffenbaugh, N.S., Field, C.B., 2013. Changes in ecologically critical terrestrial climate conditions. Science 341, 486–492.

Earle, P.S., Bowden, D.C., Guy, M., 2012. Twitter earthquake detection: earthquake monitoring in a social world. Ann. Geophys. 54.

Emergency Events Database, .. Emergency Events Database. https://www.emdat.be/database, Accessed date: August 2016.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Soc. Networks 1, 215–239.

Freeman, M., 2011. Fire, wind and water: social networks in natural disasters. Journal of Cases on Information Technology (JCIT) 13, 69–79.

Gao, H., Barbier, G., Goolsby, R., Zeng, D., 2011. Harnessing the Crowdsourcing Power of Social Media for Disaster Relief DTIC Document.

Gladwin, H., Lazo, J.K., Morrow, B.H., Peacock, W.G., Willoughby, H.E., 2007. Social science research needs for the hurricane forecast and warning system. Natural Hazards Review 8, 87–95.

Gladwin, H., Morrow, B., Lazo, J., 2013. Communication and Understanding of Hurricane Sandy Storm Surge Forecast and Warning Information Supplementary Slides.

Guan, X., Chen, C., Work, D., 2016. Tracking the Evolution of Infrastructure Systems and Mass Responses Using Publically Available Data. PLoS One 11, e0167267.

Guikema, S.D., 2009. Infrastructure design issues in disaster-prone regions. Science 323, 1302–1303.

Guy, M., Earle, P., Ostrum, C., Gruchalla, K., Horvath, S., 2010. Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. International Symposium on Intelligent Data Analysis. Springer, pp. 42–53.

Halverson, J.B., Rabenhorst, T., 2013. Hurricane Sandy: the science and impacts of a superstorm. Weatherwise 66, 14–23.

Hasan, S., Ukkusuri, S.V., 2011. A threshold model of social contagion process for evacuation decision making. Transp. Res. B Methodol. 45, 1590–1605.

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.

Hasan, S., Ukkusuri, S.V., 2015. Location contexts of user check-ins to model urban geo life-style patterns. PLoS One 10, e0124819.

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. ACM, p. 6.

Hasan, S., Ukkusuri, S.V., Zhan, X., 2016. Understanding social influence in activity-location choice and life-style patterns using geo-location data from social media. Frontiers in ICT 3, 10.

Helbing, D., 2013. Globally networked risks and how to respond. Nature 497, 51–59.

Hughes, A.L., Palen, L., 2009. Twitter adoption and use in mass convergence and emergency events. Int. J. Emerg. Manag. 6, 248–260.

Hughes, A.L., St Denis, L.A., Palen, L., Anderson, K.M., 2014. Online public communications by police & fire services during the 2012 Hurricane Sandy. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems. ACM, pp. 1505–1514.

Imran, M., Elbassuoni, S.M., Castillo, C., Diaz, F., Meier, P., 2013. Extracting Information Nuggets From Disaster-related Messages in Social Media. Proc. of ISCRAM, Baden-Baden, Germany.

Jin, P., Cebelak, M., Yang, F., Zhang, J., Walton, C., Ran, B., 2014. Location-based social networking data: exploration into use of doubly constrained gravity model for origin-destination estimation. Transportation Research Record: Journal of the Transportation Research Board 72–82.

Kaufman, S., Qing, C., Levenson, N., Hanson, M., 2012. Transportation During and After Hurricane Sandy. pp. 1–34.

Kennedy, D., 2002. Science, terrorism, and natural disasters. Science 295, 405.

Kinney, R., Crucitti, P., Albert, R., Latora, V., 2005. Modeling cascading failures in the North American power grid. The European Physical Journal B-Condensed Matter and Complex Systems 46, 101–107.

Klaus, A., Yu, S., Plenz, D., 2011. Statistical analyses support power law distributions found in neuronal avalanches. PLoS One 6, e19779.

Klomp, J., Valckx, K., 2014. Natural disasters and economic growth: a meta-analysis. Glob. Environ. Chang. 26, 183–195.

Kogan, M., Palen, L., Anderson, K.M., 2015. Think local, retweet global: retweeting by the geographically-vulnerable during Hurricane Sandy. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, pp. 981–993.

Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismail, M., Wallace, W., 2015. Actions are louder than words in social media. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. ACM, pp. 292–297.

Kryvasheyeu, Y., Chen, H., Moro, E., Van Hentenryck, P., Cebrian, M., 2015. Performance of social network sensors during Hurricane Sandy. PLoS One 10, e0117288.

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. Sci. Adv. 2, e1500779.

Kumar, S., Hu, X., Liu, H., 2014. A behavior analytics approach to identifying tweets from crisis regions. Proceedings of the 25th ACM Conference on Hypertext and Social Media. ACM, pp. 255–260.

Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., 2009. Life in the network: the coming age of computational social science. Science (New York, N.Y.) 323, 721.

Lee, J.H., Davis, A., Goulias, K., 2016a. Activity Space Estimation with Longitudinal Observations of Social Media Data, Paper Submitted for Presentation at the 95th Annual Meeting of the Transportation Research Board. Washington, DC.

Lee, J.H., Gao, S., Goulias, K.G., 2016b. Comparing the Origin-destination Matrices From Travel Demand Model and Social Media Data, Transportation Research Board 95th Annual Meeting.

Li, J., Rao, H.R., 2010. Twitter as a rapid response news service: an exploration in the context of the 2008 China earthquake. The Electronic Journal of Information Systems in Developing Countries 42.

Li, T., Ding, F., Sun, Q., Zhang, Y., Kinney, P.L., 2016. Heat stroke internet searches can be a new heatwave health warning surveillance indicator. Sci. Rep. 6.

Liang, Y., Zheng, X., Zeng, D.D., Zhou, X., Leischow, S.J., Chung, W., 2015. Characterizing social interaction in tobacco-oriented social networks: an empirical analysis. Sci. Rep. 5.

Lindell, M.K., Prater, C.S., 2007a. Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: examples from hurricane research and planning. Journal of Urban Planning and Development 133, 18–29.

Lindell, M.K., Prater, C.S., 2007b. A hurricane evacuation management decision support system (EMDSS). Nat. Hazards 40, 627–634.

Lindell, M.K., Lu, J.-C., Prater, C.S., 2005. Household decision making and evacuation in response to Hurricane Lili. Natural Hazards Review 6, 171–179.

Maghrebi, M., Abbasi, A., Rashidi, T.H., Waller, S.T., 2015. Complementing travel diary surveys with Twitter data: application of text mining techniques on activity location, type and time, 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE 208–213.

Malevergne, Y., Pisarenko, V., Sornette, D., 2005. Empirical distributions of stock returns: between the stretched exponential and the power law? Quantitative Finance 5, 379–401.

Malevergne, Y., Pisarenko, V., Sornette, D., 2009. Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal. Swiss Finance Institute Research Paper.

Meyer, R.J., Baker, E.J., Broad, K.F., Czajkowski, J., Orlove, B., 2014. The Dynamics of Hurricane Risk Perception: Real-time Evidence From the 2012 Atlantic Hurricane Season. American Meteorological Society.

Milgram, S., 1967. The small world problem. Psychol. Today 2, 60–67.

Murray, J.D., 2002. Mathematical Biology I: An Introduction, Vol. 17 of Interdisciplinary Applied Mathematics. Springer, New York, NY, USA.

Myers, S.A., Sharma, A., Gupta, P., Lin, J., 2014. Information network or social network?: the structure of the twitter follow graph. Proceedings of the 23rd International Conference on World Wide Web. ACM, pp. 493–498.

Nelson, B., 2013. Natural disasters: a calculated risk. Nature 495, 271–273.

Newman, M.E., 2003. The structure and function of complex networks. SIAM Rev. 45, 167–256.

Newman, M., 2010. Networks: An introduction. Oxford University Press.

Newman, M.E., Forrest, S., Balthrop, J., 2002. Email networks and the spread of computer viruses. Phys. Rev. E 66, 035101.

Pickard, G., Pan, W., Rahwan, I., Cebrian, M., Crane, R., Madan, A., Pentland, A., 2011. Time-critical social mobilization. Science 334, 509–512.

Power, R., Robinson, B., Colton, J., Cameron, M., 2014. Emergency situation awareness: Twitter case studies. International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries. Springer, pp. 218–231.

Press, F., Hamilton, R.M., 1999. Mitigating natural disasters. Science 284, 1927.

Sachtjen, M., Carreras, B., Lynch, V., 2000. Disturbances in a power transmission system. Phys. Rev. E 61, 4877.

Sadri, A.M., Ukkusuri, S.V., Gladwin, H., 2017. The role of social networks and information sources on hurricane evacuation decision making. nat. Hazard. Rev. 18 (3), p.04017005.

Sadri, A.M., Ukkusuri, S.V., Lee, S., Clawson, R., Aldrich, D., Nelson, M.S., Seipel, J., Kelly, D., 2018. The role of social capital, personal networks, and emergency responders in post-disaster recovery and resilience: a study of rural communities in Indiana. Nat. Hazards 90 (3), 1377–1406.

Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. Proceedings of the 19th International Conference on World Wide Web. ACM, pp. 851–860.

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., Kertesz, J., 2007. Generalizations of the clustering coefficient to weighted complex networks. Phys. Rev. E 75, 027105.

Scott, J., 2012. Social Network Analysis. Sage.

Shin, S.-Y., Seo, D.-W., An, J., Kwak, H., Kim, S.-H., Gwack, J., Jo, M.-W., 2016. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci. Rep. 6.

Skinner, J., 2013. Natural disasters and Twitter: thinking from both sides of the tweet. First Monday 18.

Sornette, D., 2009. Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton University Press.

St Denis, L.A., Palen, L., Anderson, K.M., 2014. Mastering social media: an analysis of Jefferson county's communications during the 2013 Colorado floods. 11th International ISCRAM Conference, pp. 737–746.

Starbird, K., Palen, L., 2010. Pass it on?: Retweeting in mass emergency. International Community on Information Systems for Crisis Response and Management.

Travers, J., Milgram, S., 1969. An experimental study of the small world problem. Sociometry 425–443.

Ukkusuri, S., Zhan, X., Sadri, A.M., Ye, Q., 2014. Use of social media data to explore crisis informatics: study of 2013 Oklahoma tornado. Transportation Research Record: Journal of the Transportation Research Board 110–118.

Ukkusuri, S.V., Mesa-Arango, R., Narayanan, B., Sadri, A.M., Qian, X., 2016. Evolution of the Commonwealth Trade Network. International Trade Working Paper 2016/07. Commonwealth Secretariat, London.

Van Hentenryck, P., 2013. Computational disaster management. IJCAI.

Vespignani, A., 2009. Predicting the behavior of techno-social systems. Science 325, 425–428.

Vieweg, S., Hughes, A.L., Starbird, K., Palen, L., 2010. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp. 1079–1088.

Wang, Q., Taylor, J.E., 2014. Quantifying human mobility perturbation and resilience in Hurricane Sandy. PLoS One 9, e112608.

Wang, Q., Taylor, J.E., 2015. Resilience of human mobility under the influence of typhoons. Procedia Engineering 118, 942–949.

Wang, D., Lin, Y.-R., Bagrow, J.P., 2014. Social networks in emergency response. Encyclopedia of Social Network Analysis and Mining. Springer, pp. 1904–1914.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

Watts, D., Cebrian, M., Elliot, M., 2013. Dynamics of social media. Public Response to Alerts and Warnings Using Social Media: Report of a Workshop on Current Knowledge and Research Gaps. The National Academies Press, Washington, DC.

Weng, L., Menczer, F., Ahn, Y.-Y., 2013. Virality prediction and community structure in social networks. Sci. Rep. 3.

Wood, S.A., Guerry, A.D., Silver, J.M., Lacayo, M., 2013. Using social media to quantify nature-based tourism and recreation. Sci. Rep. 3.

Yang, F., Jin, P.J., Wan, X., Li, R., Ran, B., 2014. Dynamic origin-destination travel demand estimation using location based social networking data. Transportation Research Board 93rd Annual Meeting.

Zhao, S., Zhang, K., 2016. Observing Individual Dynamic Choices of Activity Chains From Location-based Crowdsourced Data, Transportation Research Board 95th Annual Meeting.

Zhong, X., Duckham, M., Chong, D., Tolhurst, K., 2016. Real-time estimation of wildfire perimeters from curated crowdsourcing. Sci. Rep. 6.