Social Network Analysis and Mining (2018) 8:56
https://doi.org/10.1007/s13278-018-0533-y

ORIGINAL ARTICLE

Analysis of social interaction network properties and growth on Twitter

Arif Mohaimin Sadri[1] · Samiul Hasan[2] · Satish V. Ukkusuri[3] · Juan Esteban Suarez Lopez[4]

Received: 26 September 2017 / Revised: 6 July 2018 / Accepted: 21 August 2018

Abstract

The complex topology of real networks allows its actors to change their functional behavior. Network models provide better understanding of the evolutionary mechanisms being accountable for the growth of such networks by capturing the dynamics in the ways network agents interact and change their behavior. Considerable amount of research efforts is required for developing novel network modeling techniques to understand the structural properties of such networks, reproducing similar properties based on empirical evidence, and designing such networks efficiently. In this study, we first demonstrate how to construct social interaction networks using social media data and then present the key findings obtained from the network analytics. We analyze the characteristics and growth of online social interaction networks, examine the network properties and derive important insights based on the theories of network science literature. We observed that the degree distributions of such networks follow power-law that is indicative of the existence of fewer nodes in the network with higher levels of interactions, and many other nodes with less interaction. While the network elements and average user degree grow linearly each day, densities of such networks tend to become zero. Largest connected components exhibit higher connectivity (density) when compared with the whole graph. Network radius and diameter become stable over time evidencing the small-world property. We also observe increased transitivity and higher stability of the power-law exponents as the networks grow. Since the data is specific to the Purdue University community, we also observe two very big events, namely Purdue Day of Giving and Senator Bernie Sanders' visit to Purdue University as part of Indiana Primary Election 2016. The social interaction network properties that are revealed in this study can be useful in disseminating targeted information during planned special events.

Keywords Planned special events · Social media · Twitter · User mention · Social interaction · Network science

✉ Arif Mohaimin Sadri
asadri@fiu.edu

Samiul Hasan
samiul.hasan@ucf.edu

Satish V. Ukkusuri
sukkusur@purdue.edu

Juan Esteban Suarez Lopez
jesuarezl@unal.edu.co

[1] Moss School of Construction, Infrastructure and Sustainability, Florida International University, 10555 West Flagler Street, Miami, FL 33174, USA

[2] Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 12800 Pegasus Drive, Orlando, FL 32816, USA

[3] Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

[4] School of Civil Engineering, National University of Colombia, Medellín, Colombia

1 Background and motivation

Network Science is an emerging research field having multifaceted outlook in the study of large-scale real networked systems which considers both the network topology and the behavior of network actors. Complex networks with dynamic and irregular structure along with their statistical properties are the primary focus of such research efforts, i.e., the combined knowledge of network structure and behavior which is significantly distinct from the straight-forward analysis of single small graphs (Albert and Barabási 2002; Newman 2003). The prevalence of networked systems has resulted in a number of studies with applications in various domains over the last decade. Such domains include social organizations, internet, systems biology, supply chains and logistics, information and communication systems, financial markets, infrastructure systems among others (Newman 2003; Boccaletti et al. 2006; Ukkusuri et al. 2016). A number of novel

Published online: 29 August 2018

structural properties and concepts have been observed and some unifying principles along with relevant statistical distributions have been derived to characterize interdependence between the structure and function of complex real networks. For example, the existence of small-world property in many real networks suggest that most vertices in the network can be reached from every other node using relatively short paths despite large size of these networks (Milgram 1967; Travers and Milgram 1969; Watts and Strogatz 1998). On the other hand, the scale-free property suggests the existence of large hubs, i.e., a few nodes which are highly connected to many other nodes in the network and such hubs result in a network degree distributions (power-law) with highly right-skewed long tail referring to nodes with a much higher degree than most other nodes (Barabási and Albert 1999). Other properties include transitivity, network resilience, mixing patterns, network homophily or similarity, degree correlations, preferential attachment, community structure, network navigation, size of largest components among other which provide valuable insights (Newman 2003).

The coupling between network structure and behavior of network agents has initiated a number of studies related to the dynamical behavior of network actors communicating through complex network topologies. This interdependence has significant outcomes when the robustness and resilience of a real network is considered and the way networks respond to targeted failure due to external disturbances (Albert et al. 2000). Fundamental research questions, as explored in the empirical literature, can be listed as (1) how to model emergence of new innovations or ideas based on agent interactions, (2) who are the key influential players in the network and how to find them, (3) how to maximize network influence based on certain mechanism, (4) when and how networks execute contagious behavior such as disease transmission or information propagation, (5) which agents are more likely to connect and interact with agents of the same profile? A few examples of such studies may include epidemic models of disease transmission (Anderson et al. 1992; Murray 2002), email networks and computer virus transmission (Balthrop et al. 2004; Newman et al. 2002), collapse in power grid networks (Kinney et al. 2005; Sachtjen et al. 2000), disruptions of trade markets (Sornette 2009), information propagation in social networks (Coleman et al. 1966), and many others.

Information dissemination is the systematic way of distributing information and spreading awareness to every individual and systematic planning, collection, organization, and delivery technique are needed before circulating relevant information to any target audience using various media and communication means. The successful spreading of awareness to every individual in a community solely depends on an effectual information dissemination process (Cutter and Finch 2008; Helbing 2013; Vespignani 2009). The prevalence of user activities in social media (Twitter and Facebook for example) has shown its prevalence over the last decade, allowing people to be more connected than ever before worldwide. This has been possible because of the flexibility in information sharing and the way people instigate their online neighbors. The benefits to convincingly collect, analyze and use such large-scale and rich information from online information sources have also been addressed (Lazer et al. 2009). Many empirical studies have explored social media data for emergency response and disaster management (Bagrow et al. 2011; 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. In fact, one could efficiently analyze and predict real world human actions based on user activity and connectivity on social media platforms (Korolov et al. 2015; Kryvasheyeu et al. 2016). However, the purpose and context of the user activity over social media platforms may vary from one user to another user. Transportation researchers have started exploiting these large-scale data sources more extensively in recent years. Such examples include problems related to travel survey techniques (Abbasi et al. 2015; Maghrebi et al. 2015), activity-pattern modeling (Hasan and Ukkusuri 2014, 2015; Zhao and Zhang 2016), origin–destination demand estimation (Cebelak 2013; Chen and Mahmassani 2016; Jin et al. 2014; Lee et al. 2016; Yang et al. 2014) and transit planning (Collins et al. 2013). However, a few studies explored people's ego-centric offline social networks and the effects of ego-centric social ties on joint activity participation (Carrasco and Miller 2009; Sadri et al. 2015) and evacuation decision-making (Sadri et al. 2017).

Social media platforms (Twitter, Facebook and others) can be considered as appropriate means of disseminating information dynamically. Studies have found that an individual's real world actions can be inferred based on the connections and activities in social media (Korolov et al. 2015). Twitter shows both the characteristics of a social network and an informational network (Myers et al. 2014) and users can share short messages up to 140 characters along with the ability to follow other. While the information network properties of Twitter instigate information contagion globally, the social network properties allows access to geographically and personally relevant information (Kryvasheyeu et al. 2016). Because of specific features, Twitter can be particularly useful for effective information dissemination. The overwhelming usage and activity of Twitter users can be expressed by a 2013 statistic that suggests over 143K tweets per second being generated on Twitter (Krikorian 2013). User activity in social media has shown its prevalence in recent years and the world is more connected now than ever before. The ease in information sharing and the ability to instigate others have primarily contributed to such connectivity all over the world. Recent studies particularly focused on the communication patterns that evolved on Twitter from social interactions more from network science perspectives (Borondo et al. 2015; Morales et al. 2015).

However, the purpose and context of the online activity on social media platforms, such as Facebook and Twitter, may vary from user to user. For example, users' check-in activity can be referred as distinct from what users' post or share to disseminate any specific information. One specific feature of such information sharing activity is the ability of users' in mentioning (direct mentions, retweets and replies) others that they follow. Recent studies have acknowledged the potential and suitability of harnessing vast Twitter information under extreme weather events. For example, some studies focused towards examining user activities based on network properties during Hurricane Sandy (Sadri et al. 2017) and community evolution in social interaction networks before and after the 2011 earthquake and tsunami in Japan (Lu and Brelsford 2014). Some studies have offered new methodologies to infer communities, i.e., group of like-minded users based on social interactions on Twitter (Sadri et al. 2017). The role of social networks and information sources on people's decision-making during such crises has also been explored (Sadri et al. 2017).

In this study, we used Twitter REST API for the keyword 'purdue' and obtained 56,159 tweets over a month. Each tweet consisted of several words and user mentions which co-appeared with the keyword 'purdue' and resembled higher likelihood of a Twitter subscriber belonging to the Purdue University community. Next, we construct the 'purdue' specific online social interaction networks by considering the user mentions. Finally, we analyze such network characteristics, examine the properties and network growth, and derive relevant insights based on the theories of network science literature. Unlike studies that primarily examined user activities (Sadri et al. 2017) and existence of communities (Lu and Brelsford 2014) in social interaction networks, this study captures a more aggregated evolution of social network interaction dynamics on a day-to-day basis based on Twitter data specific to student-based community for about a month, In particular, this study revealed how Twitter interactions grow within a community from a single tweet by keeping track of the user mentions in it. In particular, this study makes the following contributions:

- Construction of social interaction networks based on user mentions within a given context
- Analysis of the degree distributions of such networks as they grow over time
- Experimenting network models to obtain and assess key topological parameters to identify fundamental knowledge of the growth of such networks
- Establishing how structural properties of social interactions network change and reach saturation as the network elements grow over time
- Deriving relevant insights based on the theories of network science literature

2 Network data description

We used twitter REST API to collect tweets for using a specific keyword 'purdue' which is frequently used in Twitter within the Purdue community. The dataset selection was due to its relevance in the study of information flow patterns of specific topics related to Purdue University. It is equally important to understand how special events affect this behavior over time. By running the twitter REST API between April 16, 2016 and May 16, 2016 for four consecutive weeks, we were able to obtain 56,159 tweets for the query 'purdue' after initial data cleaning for non-English tweets and common stop words. 19,532 of these tweets did not include any user mentions, however, the rest of the tweets included at least one user mention in each tweet. The frequency distribution of the top 50 words in these tweets, co-appeared with 'purdue', suggests that these top 50 most frequent words contribute up to ~25% of all the words that appeared in the collected data. A word cloud of the 100 most frequent words in the dataset is presented in Fig. 1a. Different combinations of these words constitute specific topics based on which users influence one another using the user mention feature on Twitter. These frequently appeared words also suggest the emergence of event specific topics such as *Purdue Day of Giving, Senator Bernie Sanders's* visit during *Indiana Primary* among others. Celebrity players of Purdue such as *Anthony Brown* (football), *Spike Albrecht* (basketball) and others also contributed to many topics. The differences in the amount of user mentions in the tweets over days are plotted in Fig. 1b. It can be clearly seen that the number of tweets having user mentions is almost twice as the number of tweets without mentions. These tweets primarily contribute to the formation of networks of direct social influence. In Table 1, we present the amount of user mentions as was observed in the tweets.

In Twitter, users can post tweets up to 140 characters and each data point can be stored as a tuple Tweet once collected with the following information:

*Tweet (tweet_id) = {tweet, tweet_created_at, user_id, user_screen_name, user_location, user_name, user_followers_count, user_friends_count, user_statuses_*

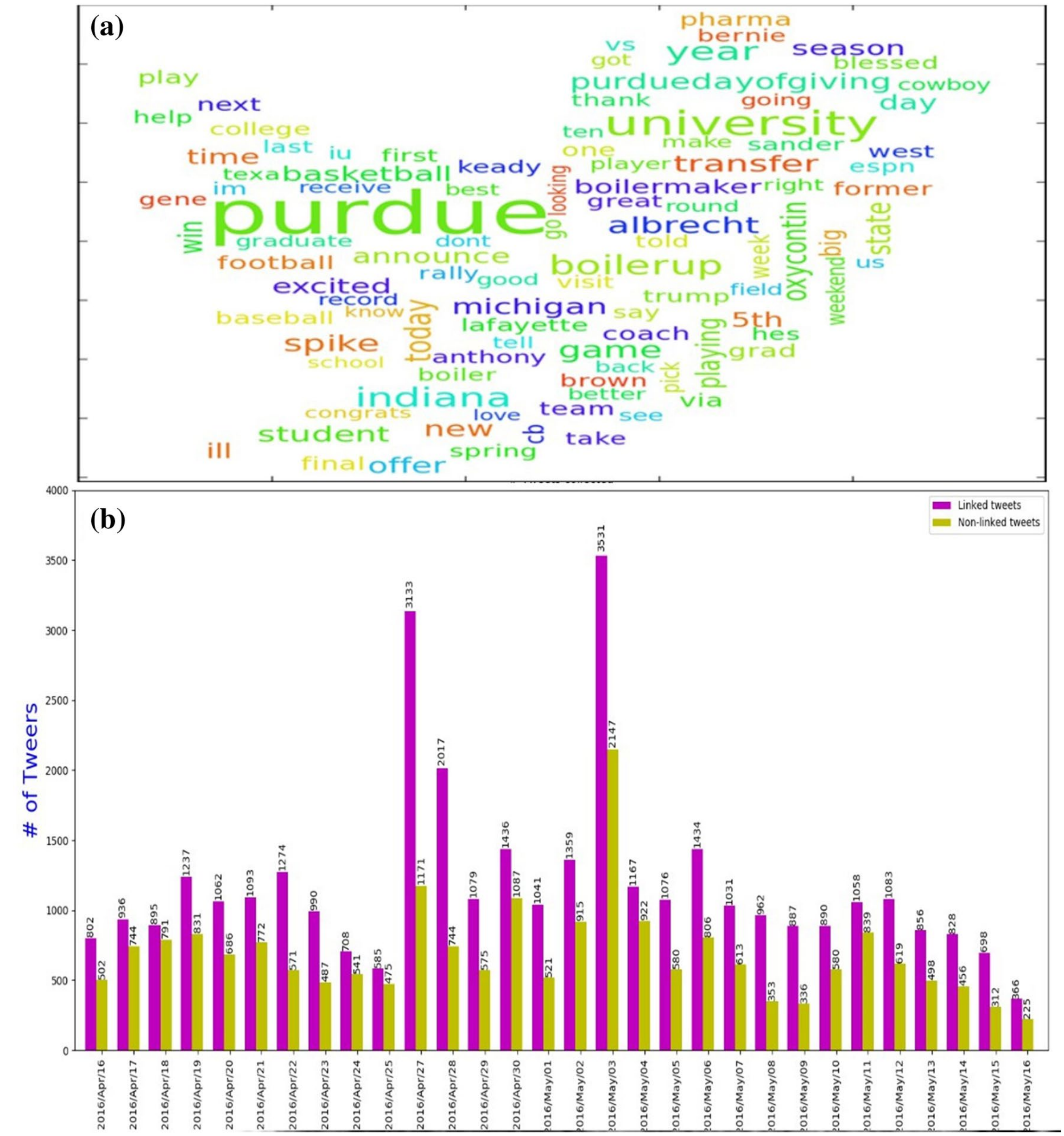

Fig. 1 Description of the Tweet database. **a** Snapshot of 100 most frequent words in the dataset, **b** Tweets collected over time (linked tweets versus Non-linked tweets)

count, user_favourites_count, user_listed_count, user_mention, tweet_retweeted, tweet_lat, tweet_lon}

For this study, we are interested in using a sub-tuple *tweet* to infer the links of direct influence that finally evolves into a highly connected network of a given context:

tweet (tweet_id) = {user_id, tweet, user_mention}

Let us consider the following three tuples from the tweets generated on Twitter on 04/28/2016 (02:45:40 + 0000), 05/02/2016 (14:45:33 + 0000) and 05/03/2016 (13:50:21 + 0000), respectively.

tweet(725516302819938305) = {709920419529281537, ***'at purdue university, we are in this campaign to win and become the democratic nominee. - bernie sanders htt…'****, [null]}*
tweet(727147016233558016) = {3239853627, ***'rt @saracohennyc at purdue university, we are in this campaign to win and become the***

Table 1 Description of the tweets and network elements

Description of the tweets	
Number of total tweets	56,159
Number of tweets without any user mentions	19,532
Number of tweets with at least one user mentions	36,627
Number of tweets only including self mentions	20,645
Number of words	3,589,732
Description of network elements	
Number of nodes (directed or undirected)	34,363
Number of links (directed)	39,709
Network density (directed)	0.00003
Number of links (undirected)	38,442
Network density (undirected)	0.00007
Number of nodes (largest connected component)	21,045
Number of links (largest connected component)	33,020
Network density (largest connected component)	0.00015
Radius (largest connected component)	9
Diameter (largest connected component)	17
Number of connected components	8348
Number of isolates	6096
Average degree (directed)	1.156
Average clustering coefficient (undirected)	0.149

democratic nominee. - bernie sanders htt...’, *[709920419529281537]}*
tweet(727495513277382656) = {325069363, ***‘rt @ bernielovesall: rt @saracohennyc at purdue university, we are in this campaign to win and become the democratic nominee. - bernie sanders htt...’***, *[3239853627, 709920419529281537]}*

Based on the above tweets, there is a directed link from user *709920419529281537* to user *3239853627* and from user *709920419529281537* to *325069363*. Please refer to Fig. 2a for the details of this network construction. The preliminary analysis of these data waves suggest the existence of 34,363 unique users and 38,442 unique undirected links (39,709 links if direction is considered) of direct influence. The network elements of the graph (constructed based on the data) are presented in Table 1.

Different network visualizations are presented from Fig. 2c–f to depict the network configurations and how the network structure appears after 4 weeks. In Fig. 2c, we present the undirected graph having 34,363 users from Twitter and 38,442 links. The network isolates without any connectivity are also shown in the periphery. This graph includes 8348 connected components and the largest connected component is presented in Fig. 2e–f. While Fig. 2f better represents the hierarchical structure of the network with the most central node in the center, Fig. 2e presents weighted edges based on the number of appearance of these links. This weighted graphs help to explain the existence of links having higher strength which also serves as an evidence of higher influence. In Fig. 2d, we present the largest hub, i.e., the most central node having the largest degree. It is highly intuitive that the network

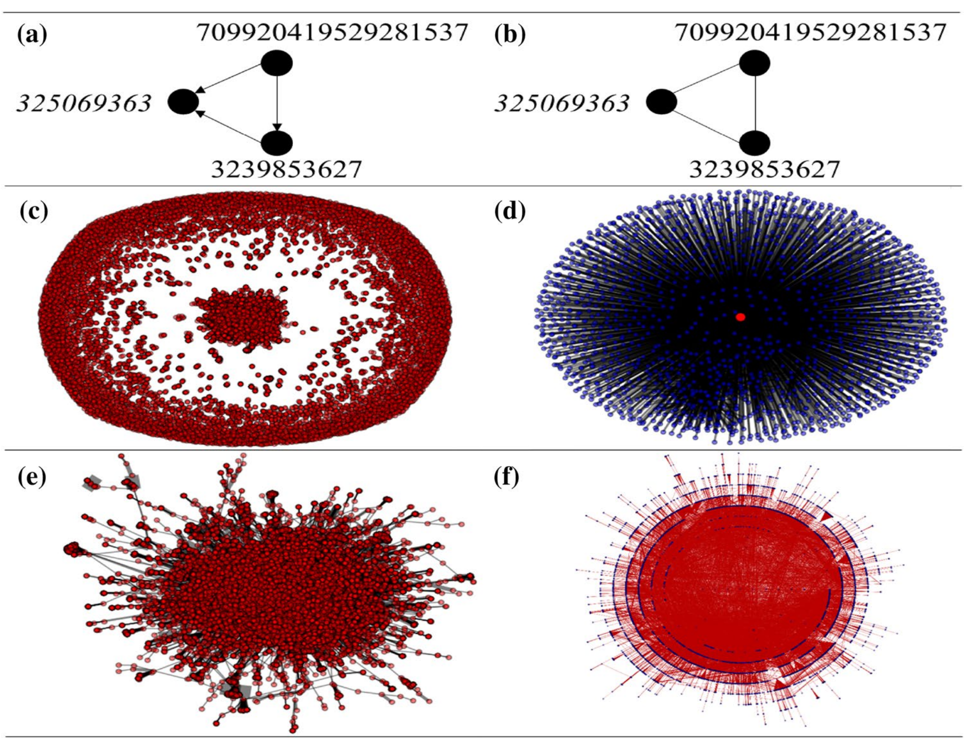

Fig. 2 Construction and visualization of social interaction network. **a** Directed graph, **b** undirected graph, **c** undirected graph visualization, **d** largest hub, **e** weighted edges of the largest connected component, **f** Circular tree visualization of the largest connected component

will be under huge disruption if such node disappears or remain active in cases.

3 Network analyses results

The user mentions, observed in the data for four consecutive weeks, construct a social interaction network that includes 34,363 nodes and 39,709 links for the directed graph and 34,363 nodes and 38,442 links for the undirected graph (Table 1; Fig. 2c). 6,096 of these nodes appeared as network isolates (nodes without connectivity) in the periphery along with 8348 connected components (Fig. 2c). 33,020 connections (links) among 21,045 users (nodes) were observed in the largest connected component of this network (Fig. 2e–f). The radius and diameter of the largest connected component were observed as 9 and 17, respectively. These are relevant to the small-world property of complex real networks that 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 has been observed in many real networks as studied in the empirical literature (Milgram 1967; Travers and Milgram 1969; Watts and Strogatz 1998). This property has significant implications in the modeling of dynamic processes occurring on real networks. For example, when effective information dissemination is considered, contagion will be faster through the network 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.

Network density, frequently used in the sociological literature (Scott 2012), equals to 0 for a graph without any link between nodes and 1 for a completely connected graph. The density of real graphs refer to the proportion of links that exist in the graph and the maximum number of possible links in the graph. For n users, the number of maximum links are $n(n-1)$ for a directed graph and $n(n-1)/2$ for a undirected graph. The densities that we observe in the social interaction network of 21,045 users are 0.00003, 0.00007 and 0.00015 for the directed, undirected and the largest connected component, respectively. This implies higher connectivity in the largest connected component, more than twice as much as in the whole network. The node *Degree* is the number of edges adjacent to that node, *In-degree* is the number of edges pointing in to the node and Out-degree is the number of edges pointing out of the node. The degree of a node (k) is the number of direct edges to other nodes in a graph from that node and the degree distribution $P(k)$ in real networks, (probability that a randomly chosen node has degree k), is significantly different from the Poisson distribution, typically assumed in the modeling of random graphs. Real networks, in fact, exhibit a power-law (or *scale-free*) degree distribution characterized by higher densities of triangles such as cliques in a social network (Barabási and Albert 1999). Such networks also demonstrate significant correlations in terms of node degrees or node-level attributes. Existence of hubs, i.e., a few nodes that are highly connected to other nodes, in the network can also be validated by the *scale-free* phenomenon. The largest hub (or ego), as was observed in our dataset, is visualized in Fig. 2d. 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 $P_{\text{degree}}(k)$ can be written as follows:

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

where, γ is some exponent and $P_{\text{degree}}(k)$ 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 that holds the same functional form (power-laws) at all scales. The power-law $P_{\text{degree}}(k)$ remains unchanged (other than a multiplicative factor) when rescaling the independent variable$k$ by satisfying:

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

The presence of hubs that are orders of magnitude larger in degree than most other nodes is a characteristic of power-law networks. The average degree of all the 34,363 users in the social interaction network is 1.156 (Table 1) and the overall degree distributions are plotted in Fig. 3a. Figure 3b presents the degree distributions that were observed each day starting from the data period of data collection. Using Alstott et al.'s python code, we obtained the best fitting to the degree distributions (Alstott et al. 2014) and the empirical data of this study fits close to being a power-law or truncated power-law distributions. The code also returns a value of x_{min} which refers to the minimal value of x at which the power-law begins to become valid. For power-law, we obtain $\gamma = 2.294$; $x_{\text{min}} = 11$ and for truncated power-law $\gamma = 2.278$; $x_{\text{min}} = 11$. Here, γ is the slope of the distribution. When γ is high, the number of nodes with high degree is smaller than the number of nodes with low degree. A low value of γ may refer to a more equal distribution, whereas higher values of γ may denote more and more unfair degree distributions. It is important to note here that the best fit power-law may only cover a portion of the distribution's tail (Alstott et al. 2014). From Fig. 3d, it appears that the data also fits close to being a log-normal distribution. However, difficulties in distinguishing the power-law from the log-normal are common and well-described, and similar issues apply to the stretched exponential and other heavy-tailed

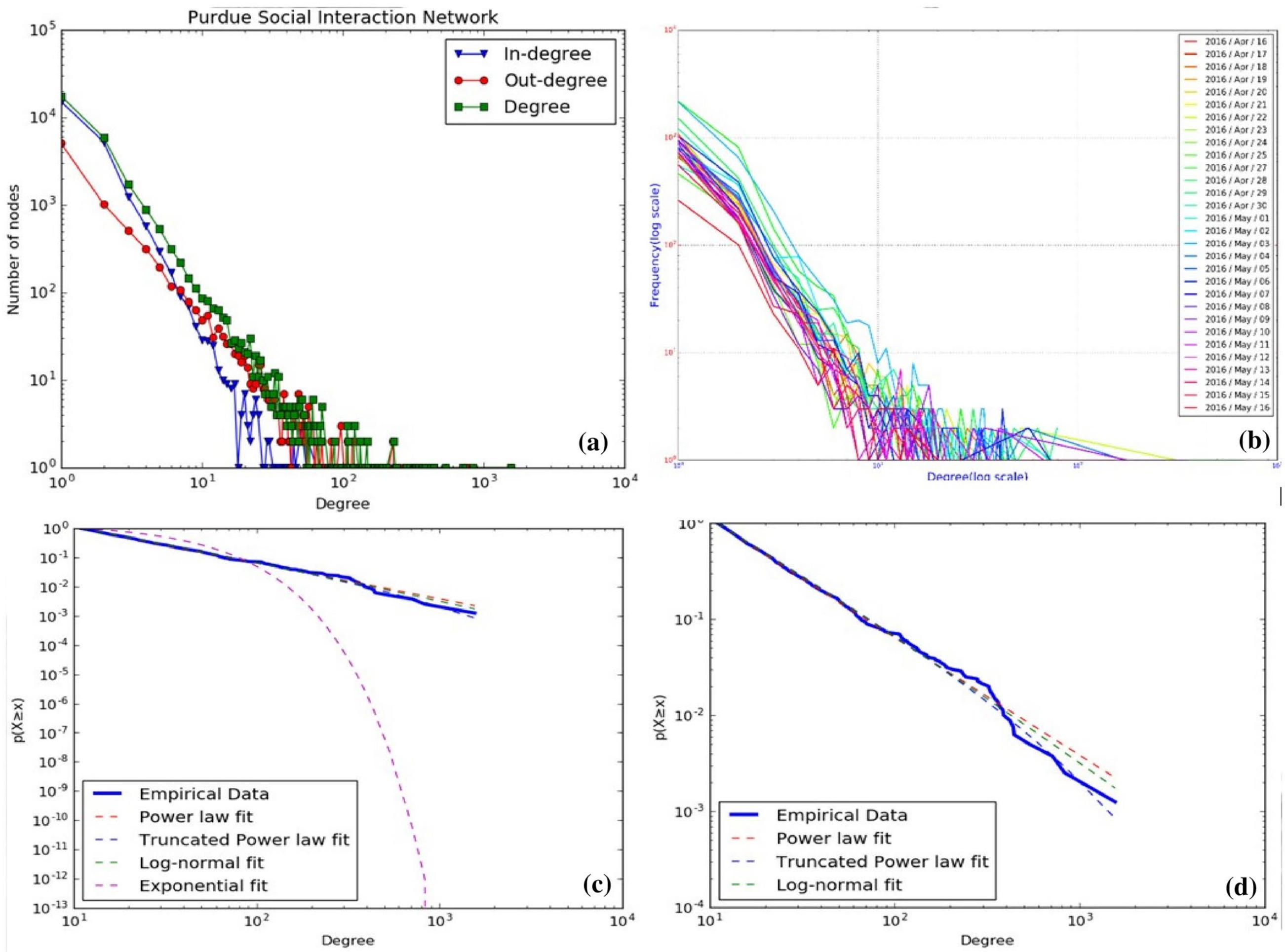

Fig. 3 Degree distributions of Purdue Twitter mention network. **a** In-degree, out-degree and degree distributions, **b** degree distributions each day, **c** comparison of data fitting with different distributions, **d** closer snapshot to power-law, truncated power-law and log-normal fitting comparisons

distributions (Malevergne et al. 2009, 2005). Our analysis on the distributions fitting are based on pairwise comparison between power-law, truncated power-law, log-normal, and exponential distributions. See Fig. 3c–d for details.

Another network property is *Transitivity* that implies the higher likelihood of any two given nodes in a network to be connected, given each of these two nodes are connected to some other node. This property refers to the fact that the friend of one's friend is likely also to be the friend of that person in case of social networks and this is a distinctive deviation from the properties of random graphs. In fact, this is indicative of heightened number of triangles (sets of three nodes each of which is connected to each of the others) that exist in real networks (Newman 2003). The existence of triangles can be quantified by *Clustering Coefficient. C*:

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

A *connected triple* refers to a single node with links running to an unordered pair of others. In case of an unweighted graph, the clustering coefficient (cc_i) of a node $i$refers to the fraction of possible triangles that exist through that node:

$$\mathrm{cc}_i = \frac{2\,T_i}{\deg_i * [\deg_i - 1]} \tag{4}$$

Here, T_i is the number of triangles that exist through node i and $\deg_i$ is the degree of node i. The average clustering coefficient in the undirected social interaction network was observed to be 0.149 (Table 1).

Turning to the results obtained from the network growth analysis, we present these results in Figs. 4 and 5. The unit of time for the analysis of network growth was set to be 24-h. We observe that the growth of network elements [nodes and links in Fig. 4a, isolates and connected components

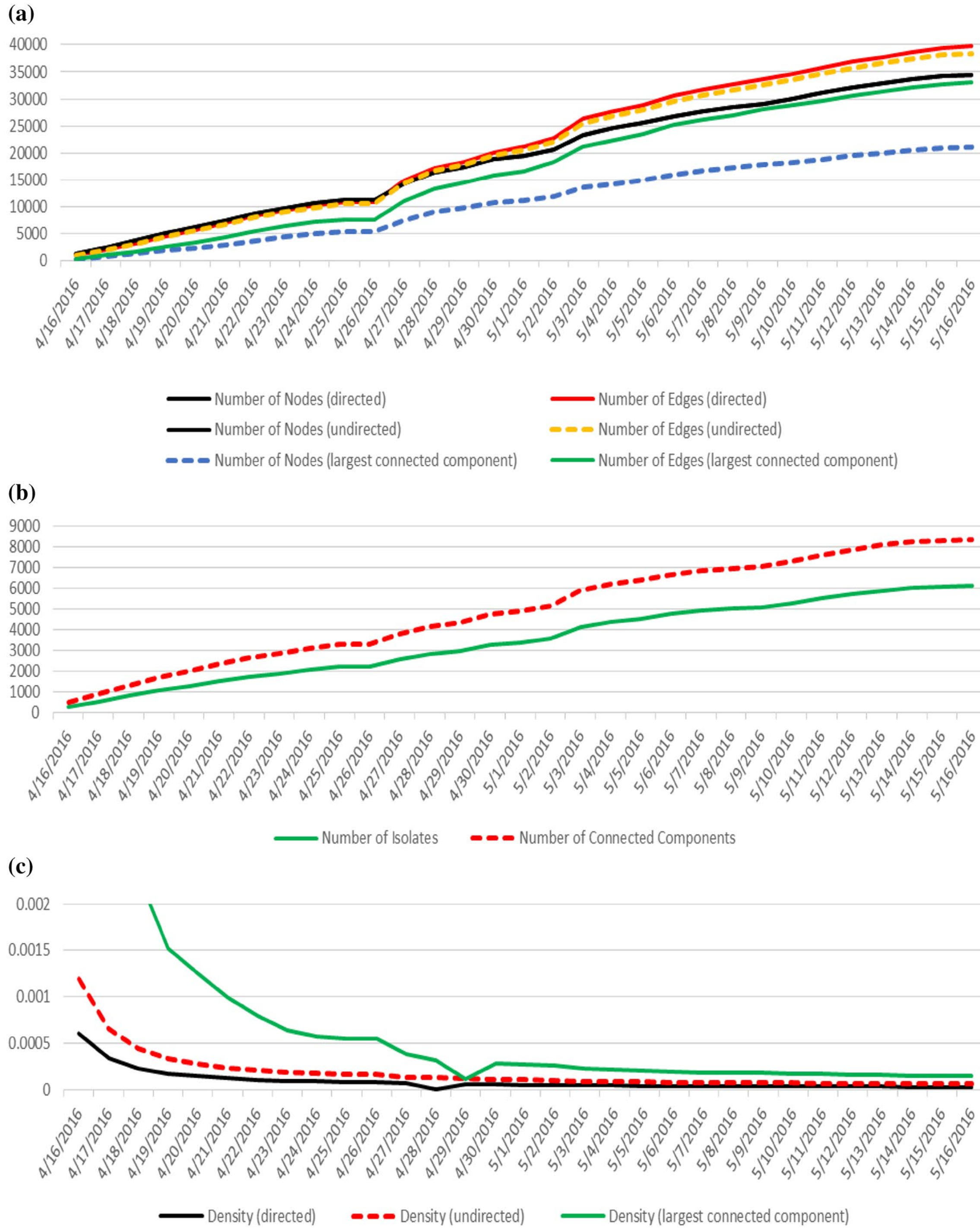

Fig. 4 Growth of Social Interaction Network over Time. **a** Cumulative amount of network elements (nodes and edges), **b** cumulative network isolates and connected generated each day, **c** change in network densities each day due to additional elements

in Fig. 4b] is almost linear over days except for the date 04/26/2016 for which we could not obtain any data. One key insight here is that the growth rate is higher for days that followed special events such as Purdue Day of Giving and Senator Bernie Sanders' visit to Purdue University during Indiana Primary 2016. While the network elements grow linearly, the densities tend to go down to zero because of less overall connectivity in the network (Fig. 4c). However, the density of the largest connected components remain slight higher over time. In addition, as the social interaction network keeps growing based on difference in user interaction for various topics, the diameter and radius keeps fluctuating

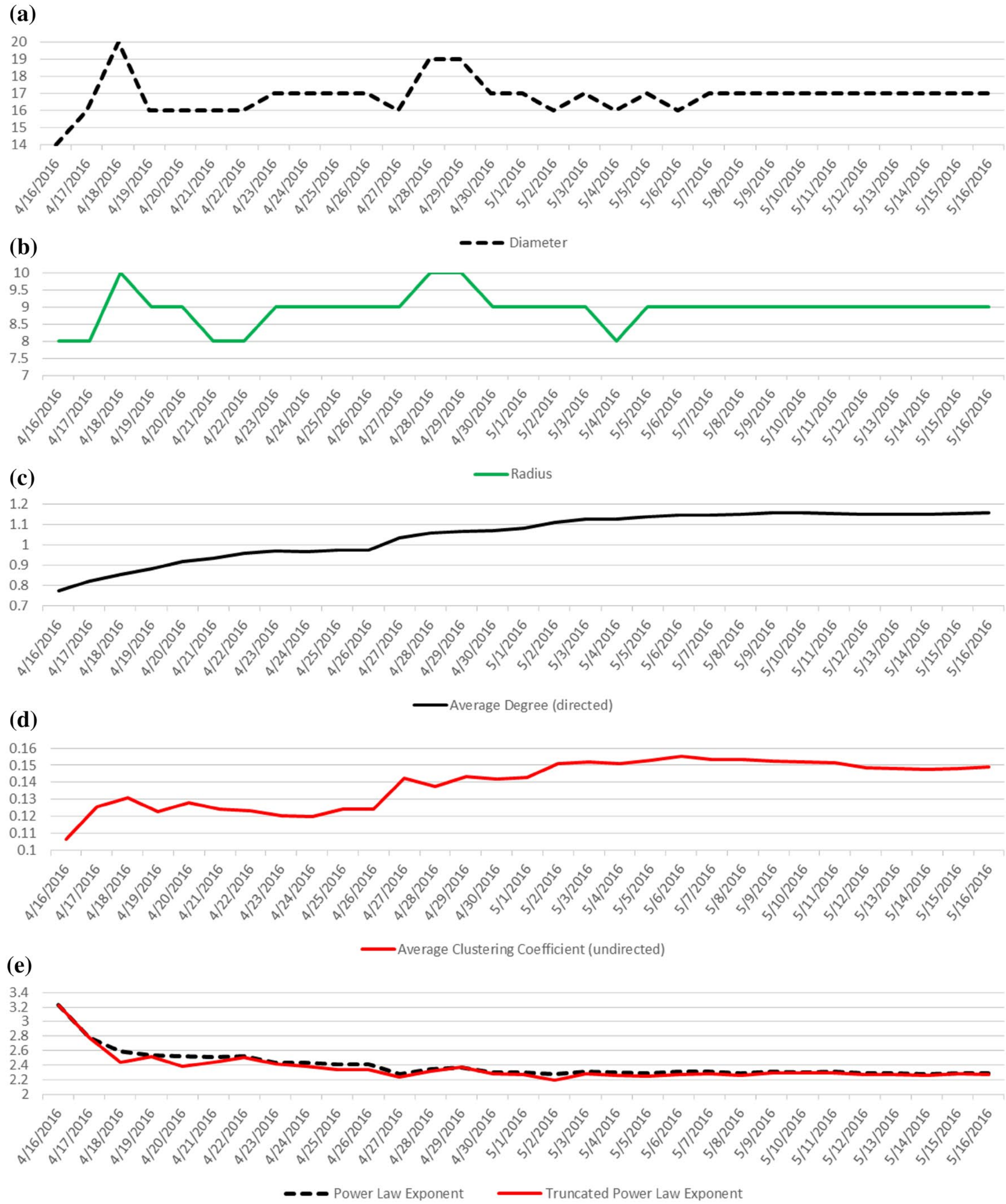

Fig. 5 Change in social interaction network properties over time. **a** Diameter, **b** radius, **c** average degree, **d** average clustering coefficient, **e** power-law and truncated power-law exponents

initially, however, becomes constant later. This is indicative of network stability when the reachability from one node to another node is considered (Fig. 5a, b). The average degree of the nodes shows similar pattern to that of the growth of network elements initially, however, becomes flat later (Fig. 5c). The network transitivity, based on average clustering coefficient, suggests that the network becomes more transitive over time, however, slight fluctuation still remains (Fig. 5d). The power-law exponents each day are presented in Fig. 5e. After reducing slightly in the initial days, they turn to becoming flat and take a value close to 2.3. This implies that the power-law property holds when social interaction network is observed over a long period of time.

Finally, we present the existence of repetitions in terms of how elements of such networks appear in the network data (Fig. 6). This is of great significance within the context of finding highly active nodes (users) in the social interaction networks along with the strength of relationships between node pairs. The relevance of considering the dynamic strength of social ties in information spreading has been duly addressed (Granovetter 1973; Miritello et al. 2011). The weighted graph, based on the number of times a link has appeared, is presented in Fig. 2e. Tob assess the commonalities of network elements (nodes and links) over time, we compute the fraction of nodes and links every day that appeared at least once in any of the previous days. From Fig. 6a, it can be seen that 65.2% of the total users (or nodes) on May 16, 2016 appeared in the data at least once. Similarly, we observe that 28.3% of the total links of interaction (undirected graph) on May 16, 2016 appeared, in any of the previous days, at least once (Fig. 6b).

4 Conclusions and key findings

Real networks having complex topologies demonstrate a unique interdependence between the structure and functional behavior. In this study, we demonstrate such interdependence by exploiting online social interaction networks based on network data obtained from Twitter. The social interaction network was formed by following the user mentions appeared in the tweets during four consecutive weeks which are specific to a university community. The network characteristics and properties have been analyzed and the network growth has been monitored over time. Key insights obtained from the network analyses are listed below:

- The network degree distributions exhibit a power-law which is indicative of the scale-free property of most real networks. This property holds for any given day as evident from the empirical data. This is indicative of the existence of fewer nodes in the network with higher levels of interactions, and many other nodes with less interactions.
- Network visualization is indicative of some nodes (users) being highly active, some links (relations) having higher strength, existence of network isolates, connected components, and hubs, i.e., nodes having reachability to many other nodes. This is also evident when the appearance of the network elements each day is compared to all previous days.
- Network elements and average user degree grow linearly each day, however, network densities tend to become zero. Largest connected components exhibit higher connectivity (density) when compared with the whole graph.
- Network radius and diameter become stable over time which suggests less variations when the reachability from one node to another node is considered. These variables are related to the small-world property.
- Increased transitivity in the growth of such networks is observed following the pattern of mean clustering coefficient. Initial fluctuations of the power-law exponents reduce as the network grows.

Planned Special Events (PSE) include sporting events, concerts, conventions and similar large events at specific venues such as stadiums and convention centers among others. PSEs possess many operational needs that can be anticipated and managed in advance because of specific locations and times of occurrence (Skolnik et al. 2008). Organizing PSEs have several challenges including parking management, crowd management, pedestrian facility design, and special facility for senior citizens and handicapped individuals, providing transit facility for captive riders among others. In addition, police enforcements often need to close several streets for security reasons, manage crowds who walk together to the location and guide motorists to specific routes who are unfamiliar with the area. Individuals attending these events travel by various travel modes, i.e., walk, private car and public transit. Since the traffic patterns of PSEs vary significantly as compared to any given weekday traffic patterns, accidents, or any other incidents, it is of great inconvenience for traffic managers, drivers or freight movers to deal with PSEs (Skolnik et al. 2008). Thus, PSEs are a major concern for traffic planners and local transportation agencies because of increased traffic demand and restricted roadway capacity causing disruptions to the regular traffic conditions (Skolnik et al. 2008; Carson and Bylsma 2003). However, this disruption and the associated operational needs can be anticipated and managed in advance (Latoski et al. 2003). Participation from key stakeholders, development and implementation of effective traffic management plan, and the flexibility to change plans to manage real-time traffic are among the key strategies to efficiently handle PSEs (Dunn 1989). Information dissemination thus constitutes an important and critical factor for the success of organizing PSEs. Despite many technical challenges to manage PSEs, the empirical literature does not provide any specific guidelines to local traffic managers and emergency response personnel to disseminate travel specific information as part of traffic management plans for PSEs. Unlike the usual communication media, social media can be used as useful means to disseminate information during any specific event such as PSEs.

The properties of social interaction networks, as observed in this study, have fundamental implications towards effective information dissemination. For example, power-law degree distributions are related to the resiliency of a communication network. The level of resilience, when random

Fig. 6 Existence of common nodes and links each day as compared to all previous days. **a** Common users, **b** common links (undirected)

nodes in the network are removed, depends solely on the way the network is formed, i.e., network topology. In case of networks having many low-degree nodes would have less disruption and higher resilience since these nodes lie on few paths between others. However, removal of hubs (high degree nodes) would cause major disruption and network agents would fail to communicate since the regular length of path will increase as a result of many disconnected pairs of nodes. For any *Planned Special Event (PSE)*, the assembling of vehicles and pedestrians in a short amount of time cause transportation and transit authorities to often encounter significant challenges in controlling the induced traffic coming from different origins before the event and departing from the event location after the event. There is hardly any specific method in the empirical literature that would allow local emergency managers or agencies to properly disseminate targeted information to any specific audience as part of traffic management procedures for PSEs. A better knowledge of social interaction network growth and properties would be worthwhile to be considered for such events. Future studies should also examine the similarity and differences between the growth patterns of social interaction networks on various topics. In addition, the differences in network properties as compared to random graphs should also be considered.

Acknowledgements The authors are grateful to National Science Foundation for the Grant CMMI-1131503 and CMMI-1520338 to support the research presented in this paper. However, the authors are solely responsible for the findings presented in this study.

Author contributions All the authors have contributed to the design of the study, conduct of the research, and writing the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no competing financial interests.

References

Abbasi A, Rashidi TH, Maghrebi M, Waller ST (eds) (2015) Utilising location based social media in travel survey methods: bringing Twitter data into the play. In: Proceedings of the 8th ACM SIGSPATIAL international workshop on location-based social networks. ACM

Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47

Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. Nature 406(6794):378–382

Alstott J, Bullmore E, Plenz D (2014) powerlaw: a Python package for analysis of heavy-tailed distributions. PloS One 9(1):e85777

Anderson RM, May RM, Anderson B (1992) Infectious diseases of humans: dynamics and control: Wiley Online Library

Bagrow JP, Wang D, Barabasi A-L (2011) Collective response of human populations to large-scale emergencies. PloS One 6(3):e17680

Balthrop J, Forrest S, Newman ME, Williamson MM (2004) Technological networks and the spread of computer viruses. Science 304(5670):527–529

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: structure and dynamics. Phys Rep 424(4):175–308

Borondo J, Morales A, Benito R, Losada J (2015) Multiple leaders on a multilayer social media. Chaos Solitons Fractals 72:90–98

Caragea C, McNeese N, Jaiswal A, Traylor G, Kim H-W, Mitra P et al (eds) (2011) Classifying text messages for the haiti earthquake. In: Proceedings of the 8th international conference on information systems for crisis response and management (ISCRAM2011). Citeseer

Carrasco J-A, Miller EJ (2009) The social dimension in action: a multilevel, personal networks model of social activity frequency between individuals. Transp Res Part A Policy Pract 43(1):90–104

Carson JL, Bylsma RG (2003) Transportation planning and management for special events

Cebelak MK (2013) Location-based social networking data: doubly-constrained gravity model origin-destination estimation of the urban travel demand for Austin, TX

Chen Y, Mahmassani HS (eds) (2016). Exploring activity and destination choice behavior in two metropolitan areas using social networking data. In: Transportation research board 95th annual meeting

Coleman JS, Katz E, Menzel H (1966) Medical innovation: a diffusion study: Bobbs-Merrill Co

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

Cutter SL, Finch C (2008) Temporal and spatial changes in social vulnerability to natural hazards. Proc Natl Acad Sci 105(7):2301–2306

Dunn W Jr (1989) Traffic management of special events: the 1986 US Open Golf Tournament. Trans Res Circ 344

Earle PS, Bowden DC, Guy M (1989) Twitter earthquake detection: earthquake monitoring in a social world. Ann Geophysics 54(6)

Freeman M (2011) Fire, wind and water: social networks in natural disasters. JCIT 13(2):69–79

Granovetter MS (1973) The strength of weak ties. Am J Sociol 1360–1380

Guy M, Earle P, Ostrum C, Gruchalla K, Horvath S (eds) (2010). Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. In: International symposium on intelligent data analysis. Springer

Hasan S, Ukkusuri SV (2014) Urban activity pattern classification using topic models from online geo-location data. Transp Res Part C Emerg Technol 44:363–381

Hasan S, Ukkusuri SV (2015) Location contexts of user check-ins to model urban geo life-style patterns. PloS One 10(5):e0124819

Helbing D (2013) Globally networked risks and how to respond. Nature 497(7447):51–59

Hughes AL, Palen L (2009) Twitter adoption and use in mass convergence and emergency events. Int J Emerg Manage 6(3–4):248–260

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. Transp Res Rec J Transp Res Board 2430:72–82

Kinney R, Crucitti P, Albert R, Latora V (2005) Modeling cascading failures in the North American power grid. Eur Phys J B Condens Matter Complex Syst 46(1):101–107

Korolov R, Peabody J, Lavoie A, Das S, Magdon-Ismail M, Wallace W (eds) (2015) Actions are louder than words in social media. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015. ACM

Krikorian R (2013) New tweets per second record, and how. Twitter Eng Blog 16

Kryvasheyeu Y, Chen H, Obradovich N, Moro E, Van Hentenryck P, Fowler J et al (2016) Rapid assessment of disaster damage using social media activity. Sci Adv 2(3):e1500779

Latoski SP, Dunn WM, Wagenblast B, Randall J, Walker MD (2003) Managing travel for planned special events: final report

Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D et al (2009) Life in the network: the coming age of computational social science. Science 323(5915):721

Lee JH, Gao S, Goulias KG (eds) (2016). Comparing the origin-destination matrices from travel demand model and social media data. In: Transportation research board 95th annual meeting

Li J, Rao HR (2010) Twitter as a rapid response news service: an exploration in the context of the 2008 China earthquake. Electron J Inf Syst Dev Countries 42

Lu X, Brelsford C (2014) Network structure and community evolution on twitter: human behavior change in response to the 2011 Japanese earthquake and tsunami. Sci Rep 4:6773

Maghrebi M, Abbasi A, Rashidi TH, Waller ST (eds) (2015) Complementing travel diary surveys with twitter data: application of text mining techniques on activity location, type and time. In: 2015 IEEE 18th international conference on intelligent transportation systems. IEEE

Malevergne Y, Pisarenko V, Sornette D (2005) Empirical distributions of stock returns: between the stretched exponential and the power law? Quant Fin 5(4):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 09–40

Milgram S (1967) The small world problem. Psychol Today 2(1):60–67

Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. Phys Rev E 83(4):045102

Morales AJ, Creixell W, Borondo J, Losada JC, Benito RM (2015) Characterizing ethnic interactions from human communication patterns in Ivory Coast. NHM 10(1):87–99

Murray JD (2002) Mathematical biology I: an introduction, vol 17 of interdisciplinary applied mathematics. Springer, New York

Myers SA, Sharma A, Gupta P, Lin J (eds) (2014) Information network or social network? The structure of the twitter follow graph. In: Proceedings of the 23rd international conference on World Wide Web. ACM

Newman ME (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256

Newman ME, Forrest S, Balthrop J (2002) Email networks and the spread of computer viruses. Phys Rev E 66(3):035101

Pickard G, Pan W, Rahwan I, Cebrian M, Crane R, Madan A et al (2011) Time-critical social mobilization. Science 334(6055):509–512

Sachtjen M, Carreras B, Lynch V (2000) Disturbances in a power transmission system. Phys Rev E 61(5):4877

Sadri AM, Lee S, Ukkusuri SV (2015) Modeling social network influence on joint trip frequency for regular activity travel decisions. Transp Res Record J Transp Res Board 2495:83–93

Sadri AM, Hasan S, Ukkusuri SV, Cebrian M (2017a) Crisis communication patterns in social media during hurricane sandy. Transp Res Record. https://doi.org/10.1177/0361198118773896

Sadri AM, Hasan S, Ukkusuri SV, Cebrian M (2017b) Understanding information spreading in social media during Hurricane Sandy: user activity and network properties. arXiv preprint arXiv:170603019

Sadri AM, Ukkusuri SV, Gladwin H (2017c) The role of social networks and information sources on hurricane evacuation decision making. Nat Hazards Rev. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000244

Sadri AM, Ukkusuri SV, Gladwin H (2017d) Modeling joint evacuation decisions in social networks: the case of Hurricane Sandy. J Choice Model 25:50–60

Sadri AM, Hasan S, Ukkusuri SV (2017e) Joint inference of user community and interest patterns in social interaction networks. arXiv preprint arXiv:170401706

Sakaki T, Okazaki M, Matsuo Y (eds) (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. ACM

Scott J (2012) Social network analysis. Sage

Skinner J (2013) Natural disasters and Twitter: thinking from both sides of the tweet. First Monday 18(9)

Skolnik J, Chami R, Walker M (2008) Planned special events—economic role and congestion effects

Sornette D (2009) Why stock markets crash: critical events in complex financial systems. Princeton University Press, Princeton

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

Ukkusuri S, Zhan X, Sadri A, Ye Q (2014) Use of social media data to explore crisis informatics: Study of 2013 Oklahoma tornado. Transport Res Rec J Transp Res Board 2459:110–118

Ukkusuri SV, Mesa-Arango R, Narayanan B, Sadri AM, Qian X (2016) Evolution of the commonwealth trade network. In: International Trade Working Paper 2016/07, Commonwealth Secretariat, London

Van Hentenryck P (ed) (2013) Computational disaster management. IJCAI

Vespignani A (2009) Predicting the behavior of techno-social systems. Science 325(5939):425–428

Wang D, Lin Y-R, Bagrow JP (2014) Social networks in emergency response. Encyclopedia of social network analysis and mining. Springer. p 1904–1914

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):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

Yang F, Jin PJ, Wan X, Li R, Ran B (eds) (2014) Dynamic origin-destination travel demand estimation using location based social networking data. In: Transportation research board 93rd annual meeting

Zhao S, Zhang K (eds) (2016) Observing individual dynamic choices of activity chains from location-based crowdsourced data. In: Transportation research board 95th annual meeting