exeley
Article | DOI: 10.21307/connections-2019.018

CONNECTIONS
Issue 1 | Vol. 40

COVID-19 Health Communication Networks on Twitter: Identifying Sources, Disseminators, and Brokers

Ian Kim* and Thomas W. Valente

Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA.

*E-mail: iank@usc.edu

Abstract

Coronavirus disease of 2019 (COVID-19)'s devastating effects on the physical and mental health of the public are unlike previous medical crises, in part because of people's collective access to communication technologies. Unfortunately, a clear understanding of the diffusion of health information on social media is lacking, which has a potentially negative impact on the effectiveness of emergency communication. This study applied social network analysis approaches to examine patterns of #COVID19 information flow on Twitter. A total of 1,404,496 publicly available tweets from 946,940 U.S. users were retrieved and analyzed. Particular attention was paid to the structures of retweet and mention networks and identification of influential users: information sources, disseminators, and brokers. Overall, COVID-19 information was not transmitted efficiently. Findings pointed to the importance of fostering connections between clusters to promote the diffusion in both networks. Lots of localized clusters limited the spread of timely information, causing difficulty in establishing any momentum in shaping urgent public actions. Rather than health and communication professionals, there was dominant involvement of non-professional users responsible for major COVID-19 information generation and dissemination, suggesting a lack of credibility and accuracy in the information. Inadequate influence of health officials and government agencies in brokering information contributed to concerns about the spread of dis/misinformation to the public. Significant differences in the type of influential users existed across roles and across networks. Conceptual and practical implications for emergency communication strategies are discussed.

Keywords

COVID, Information diffusion, Health communication, Social network analysis, Twitter.

Since the first case of Coronavirus Disease of 2019 (COVID-19) was confirmed in the United States on January 21, 2020, over 13 million people in the U.S. have confirmed cases of COVID-19. Despite multiple national- and state-wide interventions and prevention measures including banning non-essential travel and stay-at-home orders, on April 12, the USA became the nation with the most deaths globally. As of December 2, the U.S. death toll surpassed 271,000.

Providing up-to-date, accurate information, delivering key messages timely to the public, and controlling the spread of dis/misinformation can play a crucial role in managing epidemics (Homeland Security Council (US), 2006; World Health Organization, 2009; Centers for Disease Control and Prevention, 2014; Department of Homeland Security, 2018). COVID-19's devastating effects on the physical and mental health of the public are unlike previous medical crises,

in part because of people's collective access to communication technologies. COVID-19 is the first pandemic of its kind in the age of social media. The amount and nature of information available to the public has changed significantly and is constantly evolving. Unfortunately, a crucial but surprisingly understudied phenomenon is the diffusion of health information on social media (Zhou et al., 2018; Aramburu et al., 2020).

Twitter, a microblogging service, has become one of the most important sources of realtime news updates, with more than 64 million users in the U.S. (Kemp, 2020). According to a recently published Pew Research Center report, 68% of American adults get news on social media and 71% of Twitter users responded they use it to get daily news (Matsa and Shearer, 2018).

Twitter users send and receive short posts called *tweets* about any topic. Tweets can be up to 280 characters long and can include user *mentions* and *keywords*. Users can forward other users' tweets and these forwarded messages are called *retweets*. Mentions can be used with the *at* symbol "@" before a username to identify a specific user. By retweeting or mentioning, users are interacting with other users and share information in a conversation-like manner (Wang et al., 2015). The *hashtag* symbol "#" can be used before a relevant keyword to initiate conversations or contribute to discussions of existing topics by showing their tweets in Twitter search. The use of the hashtag on Twitter indicates self-association of a user with an issue (Gruzd et al., 2011; Gleason, 2013).

As users interact in Twitter space, they form connections that emerge into complex social network structures. Essentially, the connections are asymmetric, since a user who is retweeted or mentioned by another user does not necessarily have to reciprocate by retweeting or mentioning them back. Due to this asymmetry, users can re-create and reinforce traditional hierarchical network structures in Twitter by relying on just a few information sources or by choosing to limit interactions to a select group of similar others (Himelboim et al., 2017). Thus, the connections built among users are indicators of information sharing and network structures reflect patterns of information flow (Himelboim et al., 2017; Majmundar et al., 2018).

There are many studies that have examined the structure of communication networks on Twitter that provide insights about information flow during political campaigns and social movements (Himelboim et al., 2012; Ansari, 2013; Harris et al., 2014; Kruikemeier, 2014; Shin et al., 2017, 2018; Recuero et al., 2019). The patterns of communication and influential groups can vary across topics, cultures, or languages. Although few recent studies investigated Ebola information dissemination patterns (Harris et al., 2018; Liang et al., 2019), their analysis was limited to retweet network, which together with mention network can provide an understanding of information flow on Twitter (Conover et al., 2011). To the best of our knowledge, this is the first study to examine both the retweet and mention networks to understand the diffusion of health information on Twitter among Americans during a pandemic.

Structural characteristics were examined at the network level to address our overarching research question: *Is the current Twitter's COVID-19 communication network effectively leveraged to facilitate the flow of valid information during this crisis?* Information can diffuse most effectively during crises if the network is sufficiently dense with low rates of clustering (Himelboim et al., 2017). Ideally, the Twitter COVID-19 communication network would have large audience and spread information quickly. Thus, we evaluated information flow in the retweet and mention networks with particular attention paid to its connectivity, modularity and direction of information flow.

Influential COVID-19 Twitter users were identified as information sources, disseminators and brokers. Ideally, COVID-19 information sources would be medical/health professionals to emphasize credibility, disseminators would be communication/journalism professionals to maximize reach, and brokers would be public health and government officials to ensure that information is accurate and continues to flow (World Health Organization, 2009; Centers for Disease Control and Prevention, 2014). Thus, the aim of this study was to determine the characteristics of COVID-19 Twitter users by comparing the professional categorizations by their roles as sources, disseminators, or brokers in the retweet and mention networks. We hope this study will significantly contribute to public health by helping devise more effective emergency communication strategies and ultimately help mitigate the spread of disease and reduce misinformation.

Methods

Data

We retrieved all publicly available tweets and user information from April 13, 2020, 08:00:00 AM, to April 16, 2020, 07:59:59 AM, GMT (UTC +0), using the Twitter API with the query "contains: #COVID19 and country code: USA and language: English." This time period was chosen because the U.S. became the nation with the highest number of deaths due to

COVID-19 on April 12 and it was predicted the highest U.S. daily death rate would occur on April 15. We selected 8 AM (instead of midnight) as the temporal boundary between days because the number of tweets started increasing around 8 AM and reached its peak around 8 PM each day. Figure 1 shows the distribution of the tweets that used #COVID19 during the study period. The Twitter users' usernames, tweets, hashtags, retweet and mention relationships and self-descriptions were collected. We did not include replies to reduce the likelihood of repetition, losing context information, or producing unreliable data caused by Twitter's new feature, "hide reply."

Construction of retweet and mention networks

The data were converted into social network format using the R package "rtweet" (Kearney, 2019). We constructed retweet and mention networks as previously reported (Yang and Counts, 2010; Harris et al., 2014; Takeichi et al., 2015; Himelboim et al., 2017). In the retweet network, each node represents a Twitter user and a directed edge is attached from user B to user A, if user B retweets a tweet originally posted by user A. The mention network was constructed in the same manner based on @username mentioning. That is, a directed edge is constructed from B to A, if user B mentions user A in his/her tweet. The opposite directions of edges in these networks therefore represent potential pathways for information flow. Figure 2 shows (a) how we built the networks and (b) how information is spread in the networks.

These two network datasets contained a total of 1,404,496 directed relationships (ties) from 946,940 users (nodes). The R package "igraph" (Csardi and Nepusz, 2006) was used to calculate network-level and user-level metrics, to identify overall network structures and influential users and to provide insights for information flow. Analyses were conducted on the whole three-day set, separately on retweet and mention networks in order to compare them.

The networks were visualized using the library "NetworkX" (Hagberg et al., 2008) for programming language Python. In order to focus on detailed elements and to give a spatial understanding of social relations (i.e., segregation, interaction, and clustering), smaller networks were created using one-hour subsets of the data (Martin III, 2012; Moody et al., 2005). The time period of April 13, 2020, 05:00:00 PM to 05:59:59 PM, GMT (UTC +0), was chosen for the subset to display because it provided a finer representation of network structures than other time periods and it had the largest amount of information for both retweet and mention networks that our lab computers could analyze. The subset network's structure was representative of the whole network. Initial visualizations were attempted on each one-hour subset individually and results were very similar, so the finest representation was included in the current study. The coefficient of variation (CV) was calculated for each of the network measures: *CV*s for degree centrality=0.34 for the retweet and 0.37 for the mention networks across 72 one-hour subsets; *CV*s for density=0.58 for the retweet and 0.62 for the mention

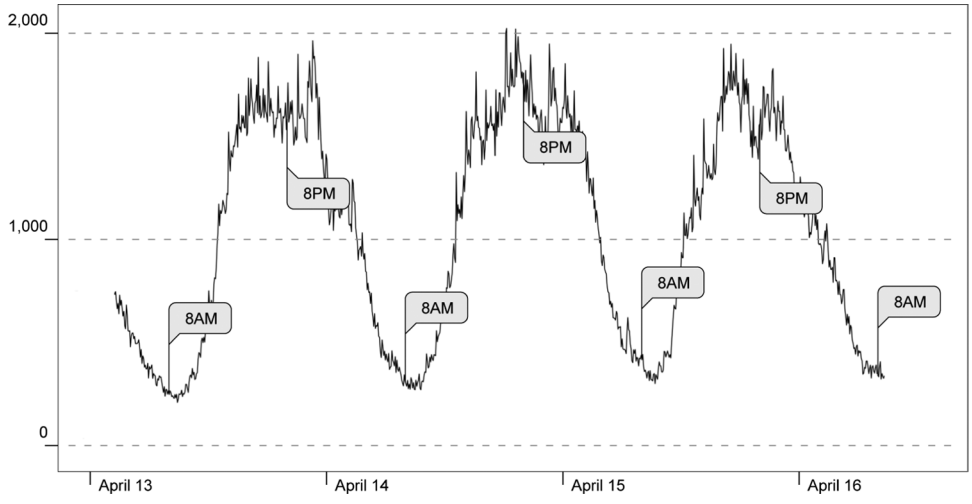
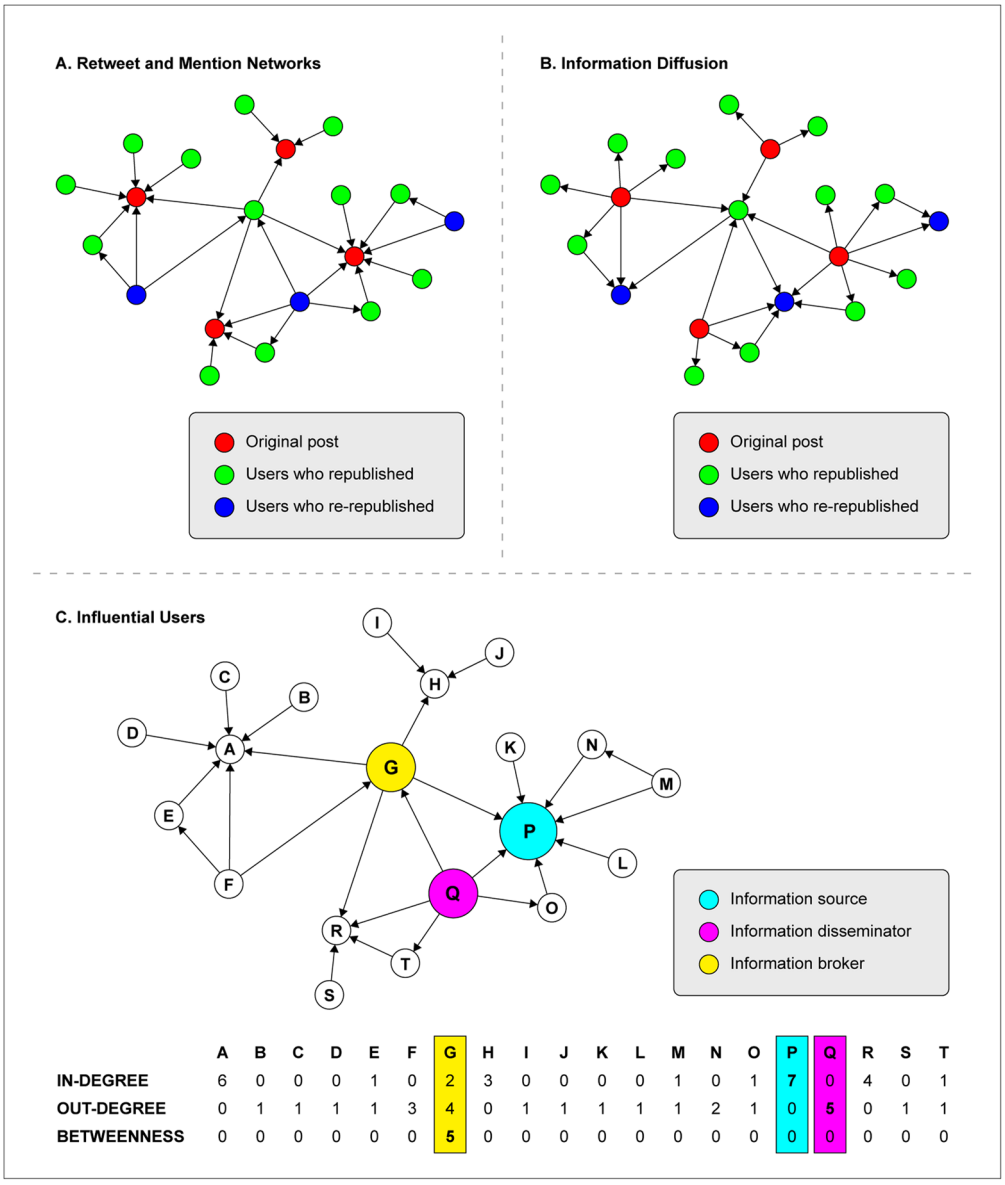


Figure 1: Volume of #COVID19 tweets from April 13, 2020, 08:00:00 AM, to April 16, 2020, 07:59:59 AM, GMT (UTC +0), with 5 minutes time intervals.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
IN-DEGREE	6	0	0	0	1	0	2	3	0	0	0	0	1	0	1	7	0	4	0	1
OUT-DEGREE	0	1	1	1	1	3	4	0	1	1	1	1	1	2	1	0	5	0	1	1
BETWEENNESS	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2: Toy networks: (a) retweet and mention networks; (b) information diffusion network; (c) influential user identification in the retweet and mention networks.

networks across 72 one-hour subsets, respectively. This indicates the network metric values for the separate one-hour slices are relatively similar.

Network level

Understanding the overall structure of a network is key for understanding how information flows among its users (Hinds and McGrath, 2006; Hossain and Kuti, 2010; Valente, 1995, 2010). Typical network level metrics are size, average path length, network diameter, rates of reciprocity and transitivity, density, as well as clustering measured as the degree of modularity and the network average clustering coefficient. Twitter users often form clusters composed of users who are more interconnected among themselves than others in the network. Within clusters, information tends to flow fast, while across clusters information flow is often restricted by limited connectivity available across clusters. We identified clusters using the Clauset–Newman–Moore algorithm to define the boundaries of information flow (Clauset et al., 2004). Modularity of each network was computed to measure the interconnectedness of

clusters using the Girvan–Newman algorithm (Girvan and Newman, 2002). Higher scores indicated that the clusters are more distinct or separated from one another (range 0=clusters completely overlap to 1=no connections between clusters). While modularity captures the extent to which clusters are distinct from one another, it is often unable to detect small clusters (Fortunato and Barthelemy, 2007; Kaalia and Rajapakse, 2019). To investigate the network in more depth, density between clusters was calculated as the sum of existing ties between two clusters divided by total possible number of ties between them (range 0=no connection to 1=complete connection).

User level

In-degree, out-degree, and betweenness centrality metrics were used to identify influential users (Freeman, 1979; Valente, 2010). Although there is no fixed ratio or standard approach to identify the number of influential users in a given network, top 10 users with highest centrality scores or more has been considered enough to provide an indication of major direction of information flow in previous studies (Anger and Kittl, 2011; Himelboim et al., 2017; Recuero et al., 2019; Giglou et al., 2020). Given the large size of our data, this study identified a total of 600 influential users from the retweet and mention networks. On Twitter, retweets and mentions are sent from one user to another. The predominant direction of such connections determines the information flow. In-degree centrality measures the number of times a user received retweets or mentions and those with high in-degree indicate the user is a major source of information for others (Yang and Counts, 2010; Morris et al., 2012; Littau and Jahng, 2016). Thus, we identified users who had the top 100 in-degree scores in each network as *information sources*. Out-degree centrality measures the number of outgoing connections a user has. If a user frequently retweets or mentions other users, the user will have high out-degree, and high out-degree will indicate the user is an initiator of large proportions of ties. Thus, we identified users who had the top 100 out-degree centrality scores in each network as *information disseminators*. Betweenness centrality measures the frequency a user lies on the shortest path between other users (Freeman, 1977, 1979). A user with high betweenness has more information passing through them and a higher number of other people depend on that user to get information, and without that user, groups of people will be much less connected. Thus, we can use this metric to find users who are communication controllers in a given network. We identified users who had the top 100 betweenness centrality scores in each network as *information brokers*. We assume that all of the connections in these networks can diffuse information equally and so centrality measures were not weighted.

During public health emergencies, health professionals have an important role to ensure the quality of shared information; likewise, the roles of communication professionals to timely disseminate the information with clear directions and of government officials to manage and maintain information flow are crucial in mitigating the effects of a pandemic (World Health Organization, 2009; Centers for Disease Control and Prevention, 2014). Interaction and cooperation between health professionals, communication professionals, and the government are critical during a pandemic (World Health Organization, 2009; Centers for Disease Control and Prevention, 2014). After identifying the information sources, disseminators and brokers, a conceptual assessment was conducted to understand the nature of influential users in the retweet and mention networks. Regarding the nature of users, we classified the users into four types, based on their self-descriptions. Healthcare providers and researchers/scientists were classified as health professionals. People who disseminate news and information to serve the public interest such as media broadcasters, journalists, and reporters were classified as communication professionals. Politicians, policy makers, and national agencies were classified as government officials. Public figures and all other ordinary individuals who are simply using Twitter to share personal views were classified as non-professionals. The user type classification results were compared across roles and across networks using Fisher's exact test.

Results

Network level

The retweet network had 646,183 ties from 438,821 users, whereas the mention network had 758,313 ties from 531,019 users. Overall, COVID-19 information was not transmitted efficiently. In both networks, information flowed in one direction; the flow was slow; both retweet and mention networks were sparse and consisted of many small clusters; the clusters were disconnected from each other; and shared information was less likely to reach the entire group. Both networks exhibited quite similar structure. Table 1 summarizes metrics from the network level analyses.

In both retweet and mention networks, low levels of mutuality of connections among users indicated

Table 1. Network metrics for the retweet and mention networks.

	Network	
	Retweet	Mention
Number of nodes	438,821	531,019
Number of edges (directed)	646,183	758,313
Diameter (largest connected component)	35	46
Average path length	12.09	16.58
Reciprocity	0.002678	0.004815
Transitivity	0.000161	0.000182
Number of clusters	12,519	28,528
Average clustering coefficient	0.012	0.008
Modularity	0.782	0.797
Density	<0.001	<0.001

the information flow is unidirectional: retweet network, reciprocity=0.268% and transitivity=0.016%; and mention network, reciprocity=0.482% and transitivity =0.018%. Both networks exhibited long average path lengths, implying information may diffuse slowly and less evenly: on average, users were separated by 12 others in the retweet network and 17 others in the mention network. Both networks were divided into a large number of clusters: 12,519 clusters in retweet network and 28,528 clusters in mention network. Information was not likely to be shared between clusters: average clustering coefficients calculated for each network were 0.012 in retweet network and 0.008 in mention network. Users had dense connections with other users within clusters but sparse connections between users in different clusters: although it was slightly lower in retweet network, both networks revealed high modularity with scores of 0.782 in retweet network and 0.797 in mention network. Both retweet and mention networks showed very low density: density scores were 0.0000034 in retweet network and 0.0000027 in mention network.

User level

Degree analyses revealed that a very small number of users determined the major COVID-19 information flow in both retweet and mention networks. The degree distributions in both networks tended to be scale-free, suggesting a hierarchical structure.

The in-degree values of all users in the retweet network ranged between 0 and 11,954 (*N*=438,821, *M*=1.47, *Med*=0), the out-degree values between 0 and 158 (*N*=438,821, *M*=1.47, *Med*=1), and the betweenness values between 0 and 43,409,213 (*N*=438,821, *M*=2,894.66, *Med*=0). In the mention network, the in-degree values of all users ranged between 0 and 11,608 (*N*=531,019, *M*=1.43, *Med*=0), the out-degree values between 0 and 187 (*N*=531,019, *M*=1.43, *Med*=1), and the betweenness values between 0 and 215,538,020 (*N*=531,019, *M*=16,452.62, *Med*=0). All degree distributions were highly right-skewed: retweet network, skewness and kurtosis scores were 132.10 and 21,154.49 for in-degree, 12.43 and 491.54 for out-degree, and 123.89 and 19,750.70 for betweenness; mention network, 140.25 and 24,930.94 for in-degree, 13.22 and 537.58 for out-degree, and 125.39 and 17,758.65 for betweenness.

The in-degree of the identified information sources (top 100) in the retweet network was between 705 and 11,954 (*N*=100, *M*=2,681, *Med*=1,506), the out-degree of the identified information disseminators was between 32 and 158 (*N*=100, *M*=47, *Med*=38), and the betweenness of the identified information brokers was between 2,728,657 and 43,309,213 (*N*=100, *M*=9,433,471, *Med*=6,886,086). In the mention network, the in-degree of the identified information sources was between 749 and 11,608 (*N*=100, *M*=2,560, *Med*=1,815), the out-degree of the identified information disseminators was between 39 and 187 (*N*=100, *M*=62, *Med*=52), and the betweenness of the identified information brokers was between 15,904,090 and 215,538,020 (*N*=100, *M*=67,239,434, *Med*=40,851,672). Table 2 compares the summary statistics of the degree distribution of influential users and of all users.

Both networks followed a power-law degree distribution, providing evidence of scale-free, hierarchical structures: in-degree α=0.957, R^2=0.694, p<0.001 and out-degree α=1.860, R^2=0.961, p<0.001 were calculated in the retweet network; in-degree α=1.019, R^2=0.704, p<0.001 and out-degree α=1.980, R^2=0.964, p<0.001 were calculated in the mention network. Figure 3 shows the scale-free in-and out-degree distributions on a log–log scale with the raw score distributions on a histogram.

The user type classification results revealed that, in both networks, the major COVID-19 information being shared among Twitter users was primarily authored by non-professionals and government officials; the information was primarily disseminated

Table 2. Centrality statistics for influential and all other users in both networks.

	Retweet		Mention	
	Influential users	All users	Influential users	All users
In-degree	*N=100*	*N=438,821*	*N=100*	*N=531,019*
Mean	2,681	1.47	2,560	1.43
SD	2,689	58.23	2,364	48.13
Median	1,506	0	1,815	0
Min.	705	0	749	0
Max.	11,954	11,954	11,608	11,608
Skewness		132.10		140.25
Kurtosis		21,154.49		24,930.94
Out-degree	*N=100*	*N=438,821*	*N=100*	*N=531,019*
Mean	47	1.47	62	1.43
SD	23	1.78	29	2.09
Median	38	1	52	1
Min.	32	0	39	0
Max.	158	158	187	187
Skewness		12.43		13.22
Kurtosis		491.54		537.58
Betweenness	*N=100*	*N=438,821*	*N=100*	*N=531,019*
Mean	9,433,471	2,894.66	67,239,434	16,452.62
SD	7,976,886	188,252.40	58,763,864	1,230,822
Median	6,886,086	0	40,851,672	0
Min.	2,728,657	0	15,904,090	0
Max.	43,309,213	43,309,213	215,538,020	215,538,020
Skewness		123.89		125.39
Kurtosis		19,750.70		17,758.65

by non-professionals; and health professionals played a major role in brokering information. The classified types of influential users in different roles in each network were all statistically significantly different from one another (all *p*s<0.001). Significant difference across networks was observed in the composition of the identified information brokers at α=0.10: Brokers in the retweet network were most frequently healthcare providers and ordinary citizens, with a near absence of government officials whereas brokers in the mention network were most often research scientists followed by healthcare workers. Table 3 summarizes the results of user level analyses. Table 4 shows the *p*-values obtained from user type composition comparison across roles and across networks using Fisher's exact test.

Retweet network

Information sources, the top 100 on in-degree, were almost evenly divided among the four user types: health professionals, 20%; communication professionals, 16%; government officials, 28%; and non-professionals, 36%. In contrast, information disseminators, the top 100 on out-degree, were predominately non-professionals, 76% (with 95% of them being ordinary people); and a handful of communication professionals, 18%. Information

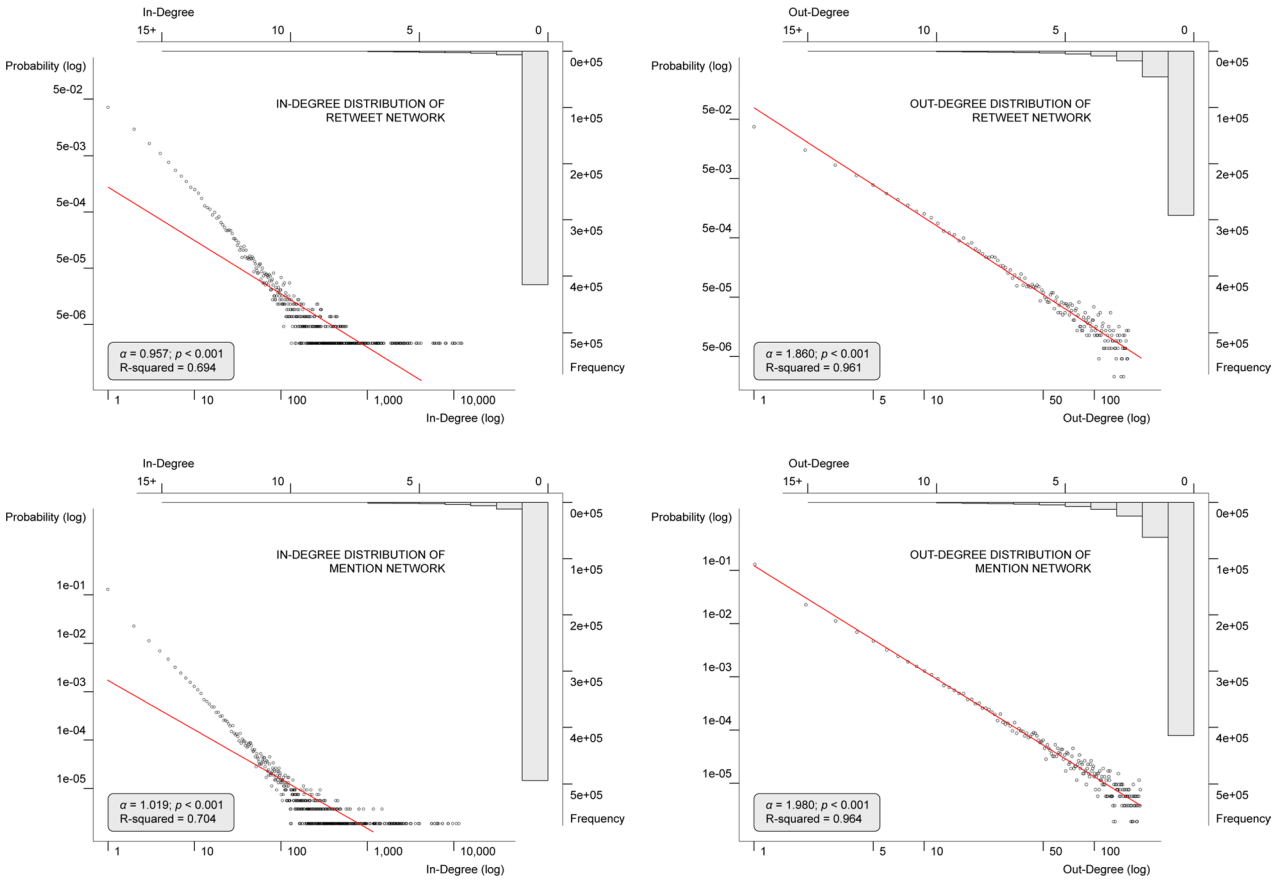


Figure 3: Scale free in-degree and out-degree distributions on a log-log scale for retweet and mention networks.
Note: Users with a degree score >15 are not shown in histograms.

brokers, the top 100 on betweenness, were predominately health professionals, 48% (with most being healthcare providers, 60%); and non-professionals being most of the remainder, 27%.

Mention network

The mention network followed a similar pattern with information sources being almost evenly divided among the four user types: health professionals, 19%; communication professionals, 18%; government officials, 34%; and non-professionals, 29%. Information disseminators, as in the retweet network, were predominately non-professionals, 69% (with 93% of them being ordinary people); and a handful of communication professionals, 17%. Information brokers were predominately health professionals, 57%, although in this case these health professionals were more likely to be researchers/scientists (61%); and government (16%) and communication professionals (15%) primarily the remainder.

Visualization

The one-hour subset data for the retweet network visualization consisted of 14,255 ties from 15,907 users. The subset data for the mention network visualization consisted of 16,379 ties from 19,386 users. Figure 4 visually depicts the structures and information flow of retweet network and mention network.

The size and color of the nodes were made proportional to the unweighted in-degree centrality score of each user. The ties between users represented the information exchange links between the users. Directions of ties were ignored. Attention was focused on the overall degree distribution and connectivity between high degree users (information sources) and lower degree users to help reveal the overall network structure and information flow. Spatialization was used to draw nodes with more ties to more central positions.

In both networks, a hierarchical structure was apparent and information flow was concentrated

Table 3. Occupations of influential users in retweet and mention networks.

	Information sources (*N*=100)			Information disseminators (*N*=100)			Information brokers (*N*=100)		
Retweet									
Health professionals	20	Care providers	16 (80.0%)	3	Care providers	1 (33.3%)	48	Care providers	29 (60.4%)
		Researchers/Scientists	4 (20.0%)		Researchers/Scientists	2 (66.7%)		Researchers/Scientists	19 (39.6%)
Communication professionals	16	Media broadcasters	3 (18.8%)	18	Media broadcasters	13 (72.2%)	13	Media broadcasters	4 (30.8%)
		Journalists/Reporters	13 (81.2%)		Journalists/Reporters	5 (27.8%)		Journalists/Reporters	9 (69.2%)
Government official	28	Politicians/Policy makers	23 (82.1%)	3	Politicians/Policy makers	1 (33.3%)	12	Politicians/Policy makers	6 (50.0%)
		National agencies	5 (17.9%)		National agencies	2 (66.7%)		National agencies	6 (50.0%)
Non-professionals	36	Public figures	29 (80.6%)	76	Public figures	4 (5.3%)	27	Public figures	6 (22.2%)
		Ordinary Individuals	7 (19.4%)		Ordinary Individuals	72 (94.7%)		Ordinary Individuals	21 (77.8%)
Mention									
Health professionals	19	Care providers	16 (84.2%)	5	Care providers	2 (40.0%)	57	Care providers	22 (38.6%)
		Researchers/Scientists	3 (15.8%)		Researchers/Scientists	3 (60.0%)		Researchers/Scientists	35 (61.4%)
Communication professionals	18	Media broadcasters	4 (22.2%)	17	Media broadcasters	13 (76.5%)	15	Media broadcasters	6 (40.0%)
		Journalists/Reporters	14 (77.8%)		Journalists/Reporters	4 (23.5%)		Journalists/Reporters	9 (60.0%)
Government official	34	Politicians/Policy makers	29 (85.3%)	9	Politicians/Policy makers	4 (44.4%)	16	Politicians/Policy makers	6 (37.5%)
		National agencies	5 (14.7%)		National agencies	5 (55.6%)		National agencies	10 (62.5%)
Non-professionals	29	Public figures	25 (86.2%)	69	Public figures	5 (7.2%)	12	Public figures	8 (66.7%)
		Ordinary individuals	4 (13.8%)		Ordinary individuals	64 (92.8%)		Ordinary individuals	4 (33.3%)

Table 4. *P*-values from Fisher's exact test comparing occupations across user roles and between networks.

	Retweet			Mention			Retweet vs. mention		
	1	2	3	1	2	3	1	2	3
1. Information sources		<0.001	<0.001		<0.001	<0.001	0.6909		
2. Information disseminators			<0.001			<0.001		0.2707	
3. Information brokers									0.0630

at the center where influential users are located. A significant portion of users in both networks were connected to only a few others, whereas a few users had a huge proportion of connections. Both networks exhibited a large core cluster, comprised of a small number of high degree users – represented by bigger and brighter nodes in the figure – surrounded by a large number of less influential users and small clusters. In both networks, information brokers played a central role in information diffusion; connections between more influential users and less influential users were mediated by others or clusters. In the retweet network, dense interconnections among influential users, connecting each of their clusters with another, were observed.

Implications

Despite Twitter's reputation as an effective medium to connect people and facilitate public communication, the topic of COVID-19 did not bring its users together. Both the retweet and mention networks were sparsely connected, exhibiting a large number of small distinct clusters. A study from Kaur and Singh (2016) reported that disconnected networks often result from distrust in information sources. Consistent with their finding, more than half of the COVID-19 information was generated by non-professional users, increasing the likelihood of encountering false information and thereby potentially spreading misinformation.

Moreover, dominant involvement of non-professional users was observed in the information dissemination process. In both the retweet and mention networks, communication professionals were only marginally involved and there were almost no health professionals among the disseminators. Since publicly shared information has a direct impact on the development of public behaviors, it is very important to consider the type of people who act as information disseminators during medical crises (Hilton and Hunt, 2011; Staniland and Smith, 2013). Findings by Keshvari et al. (2018) warned about biased and misleading content that ordinary people, who are not trained to objectively perceive risks and benefits, disseminate with personal speculations and interpretations during epidemics. Communication professionals, on the other hand, are trained to investigate all possible aspects and implications of information before promoting the information. In this process, communication professionals are often dependent on health professionals to substantiate facts and provide balance by ensuring pluralistic aspects and implications of the pandemic (Ahlmén-Laiho et al., 2014). Increasing willingness on the part of communication professionals to disseminate accurate information and to cooperate with health professionals, may be critical to control the spread of dis/misinformation and prevent public confusion.

In both the retweet and mention networks, information flow was highly concentrated within a core cluster, comprised of a few influential users and their own clusters; information flow to the rest of the network (the other clusters) was severely restricted due to the limited connectivity. This suggests that the networks facilitate the diffusion of COVID-19 information if brokers integrate with their communities and clusters. In the context of social media communications, the limited connectivity between clusters means that networks would break into isolated components, separated by redundant and unnecessary information and that information will, more often than not, be trapped within its own cluster. Brokers, on the other hand, create paths for information diffusion and make global information

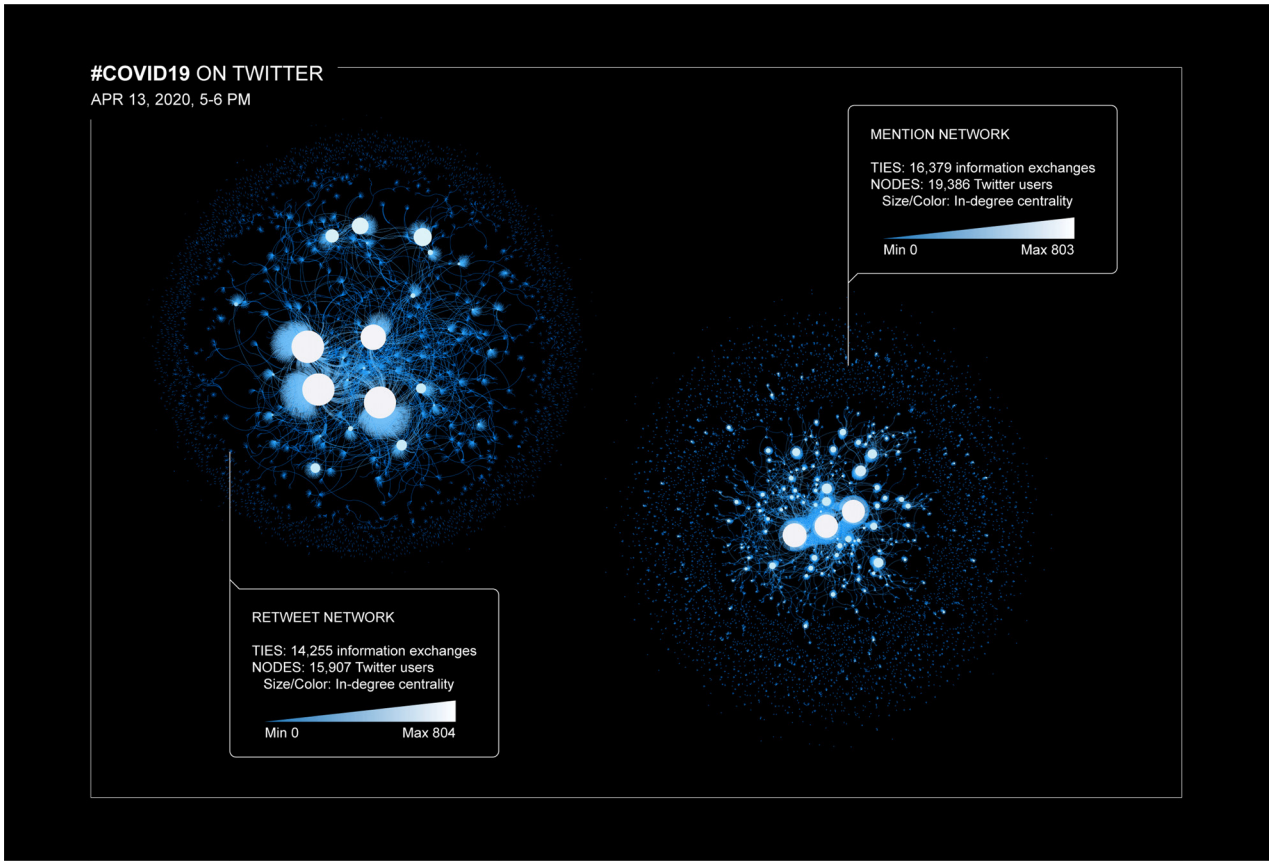


Figure 4: Graphs of the #COVID19 retweet and mention networks (April 13, 2020, 5–6 PM, GMT (UTC +0)).

flow easier to attain, if and when they are activated (González-Bailón and Wang, 2016). In a pandemic, a balanced approach to centralized control and management of information is critical in helping public audiences understand the threat and what actions should be taken (Homeland Security Council (US), 2006; Department of Homeland Security, 2018). A proper course of action as information brokers must be taken by government officials to be complete, valid and reliable (Homeland Security Council (US), 2006; Department of Homeland Security, 2018). Unfortunately, however, that was not happening in the current Twitter's COVID-19 communication network. Neither the retweet network nor the mention network showed enough influence of government officials as information brokers; in both networks, information flow was primarily maintained by health professionals. Developing social media communication guidelines for officials and national agencies that offer a starting point to foster connections and training to control or promote information flow may help ensure effective information flow and make necessary information timely and accessible to those who need it in the process of emergency response.

Both the retweet and mention networks exhibited a scale-free hierarchical structure, with unidirectional information flow. Due to preferential attachment a small number of influential users get, such network structures can be much more effective at rapid information diffusion for timely response and national solidarity during crises (Himelboim et al., 2017; De Brún and McAuliffe, 2018); because a small number of influential users can command a large and disproportionate number of other users and those users then will affect all the other users in their local network, a whole subsystem can be covered in just a few steps, making it relatively easier to keep everyone informed of relevant information such as risks and action items. At the same time, however, such network structures can also be vulnerable to false information and its diffusion can be easily distorted by just one or a few influential users' absence in the network (Lossio-Ventura and Alatrista-Salas, 2017; De Brún and McAuliffe, 2018); for instance, if one or two

influential users were removed or left the network, it would leave a major gap in support for most users thus interrupting information flow; similarly, a single piece of misinformation can be a risk factor for the entire system because of the fast nature of information dissemination. Monitoring information flow and ensuring that the public can rely on a consistently valid source of information via controlled channels at all stages of a pandemic communication planning may help the emergency communication network be more resilient and stable.

The visualization results suggested that influential users in the retweet and mention networks may have different reasons to engage in COVID-19 communication. Different interaction patterns and preferences in interaction form in Twitter networks have been previously shown to result in part from differences in the type of messages, which may reflect the reasons users engage in the communication (Conover et al., 2011; Himelboim et al., 2017). Conover et al. (2011) found that, in Twitter's political discourse where the retweet network was highly polarized while the mention network was not, users tended to retweet other users whom they agreed with politically, while they interacted with users whom they disagreed with more frequently using mentions to argue or share their views. COVID-19's retweet and mention networks did not exhibit the same connectivity among users. Interacting closely, influential users in the retweet network shared information with each other, and the interactions among influential users facilitated less influential users' access to information by connecting each of their clusters in the network. In contrast, the absence of interactions among influential users in the mention network led to the more limited information flow across clusters. Studies are needed to investigate whether and how differences in information flow tendencies in health communication represent differences as a function of information type.

Limitations

This study has some noteworthy limitations. Data collection was restricted to English messages, which may limit generalizability to other languages. The study was unable to access private networks – only publicly available tweets were retrieved for the analyses. Although a majority of Twitter users (87%) reported they keep their accounts public (Wojcik and Hughes, 2019), the findings may not reflect the characteristics and attitudes of private users. Many additional aspects of information diffusion regarding the topic of COVID-19 were not captured by the indicators of information sharing – retweets and mentions. For example, the current study did not include followers-followees structure since it has been reported that influential users are those who have an active audience who mentions or retweets the users, instead of the large number of followers (Cha et al., 2010) and the number of followers/followees does not fully explain users' actual activities (Hamzehei et al., 2017); however, it may be possible that the structure explains other aspects such as the impact of the information shared. The current version of the Twitter API does not store users who retweet retweeters. A prior study on information spread on the retweet network connection identified that most (91%) of retweets are directly retweeted from the initial message (Liang et al., 2019). However, the unavailability of the full content record may prevent us from further knowing the pattern of information diffusion among intermediate retweeters. There are no comparable analyses to determine cut-off values of network indices to be high or low. Thus, our only basis was our own interpretation of the data. Social networks are often only weakly scale-free even in cases where the power-law distribution is observed (Broido and Clauset, 2019). Future research should investigate the robustness of the scale-free structures and interpretability of power-law distribution. Drawing inferences solely based on a visual inspection requires further statistical confirmation.

Conclusion

This study examined the COVID-19 communication network on Twitter to provide insights about health information flow among Americans during a pandemic. Structural characteristics of retweet and mention networks were quantified and described with different metrics (size, density, connectivity, modularity). Influential users (information sources, disseminators, brokers) in each network were identified and the nature of the influential users were conceptually assessed. Results showed that in both retweet and mention networks, the topic of COVID-19 created large fragmented Twitter populations into multiple communication channels, each with its own audience and information sources. The study also found the absence of reliable sources, disseminators that can provide timely, accurate information, and proper management of information flow. These results have implications for understanding and predicting information diffusion in urgent public health communication. Overall, the findings emphasized the importance of connecting users to the essential resources and distinguishing credible information among a huge amount of information being shared. As social media becomes a more

heavily used news source, the effectiveness of crisis management depends more on the type of information shared among its users and the user reachability in the network. Our work opens several new questions about the underlying structures of social media communication network. Future studies may expand this research, exploring how user clusters are formed and examining how relationships between information type and degree of influence differ by cluster or change over time.

References

Ahlmén-Laiho, U., Suominen, S., Järvi, U. and Tuominen, R. 2014. "Finnish health journalists' perceptions of collaborating with medical professionals", *International Conference on Well-Being in the Information Society,* Springer, Cham, pp. 1–15.

Anger, I. and Kittl, C. 2011. "Measuring influence on Twitter", *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, pp. 1–4.

Ansari, A. 2013. "Green's art: new media aesthetics in pre-and post-election events in Iran", *Proceedings of the 19th International Symposium of Electronic Art* edited by K. Cleland, L. Fisher, and R. Harley, ISEA International, the Australian Network for Art & Technology, and the University of Sydney, Sydney.

Aramburu, M. J., Berlanga, R. and Lanza, I. 2020. Social media multidimensional analysis for intelligent health surveillance. *International Journal of Environmental Research and Public Health* 17: 2289.

Broido, A. D. and Clauset, A. 2019. Scale-free networks are rare. *Nature Communications* 10: 1–10.

Centers for Disease Control and Prevention. 2014. *Crisis and Emergency Risk Communication (CERC) Manual*, Centers for Disease Control and Prevention, Atlanta, available at: https://emergency.cdc.gov/cerc/manual/index.asp.

Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, P. K. 2010. Measuring user influence in Twitter: the million follower fallacy. *Icwsm*, 10: 30.

Clauset, A., Newman, M. E. and Moore, C. 2004. Finding community structure in very large networks. *Physical Review E* 70: 066111, doi: 10.1103/PhysRevE.70.066111.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. and Flammini, A. 2011. Political polarization on twitter. Fifth international AAAI Conference on Weblogs and Social Media.

Csardi, G. and Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal Complex Systems* 1695: 1–9.

De Brún, A. and McAuliffe, E. 2018. Social network analysis as a methodological approach to explore health systems: a case study exploring support among senior managers/executives in a hospital network. *International Journal of Environmental Research and Public Health* 15: 511.

Department of Homeland Security 2018. *Countering False Information on Social Media in Disasters and Emergencies: Social Media Working Group for Emergency Services and Disaster Management*.

Fortunato, S. and Barthelemy, M. 2007. Resolution limit in community detection. *Proceedings of the National Academy of Sciences* 104: 36–41, available at: https://doi.org/10.1073/pnas.0605965104.

Freeman, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry* 40: 35–41, available at: https://doi.org/10.2307/3033543.

Freeman, L. C. 1979. Centrality in social networks: conceptual clarification. *Social Networks* 1: 215–239.

Giglou, R. I., d'Haenens, L. and Van Gorp, B. 2020. "Identifying influential users in Twitter networks of the Turkish Diaspora in Belgium, the Netherlands, and Germany", *Handbook of Research on Politics in the Computer Age*, IGI Global, pp. 235–263.

Girvan, M. and Newman, M. E. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99: 7821–7826.

Gleason, B. 2013. #Occupy Wall Street: exploring informal learning about a social movement on Twitter. *American Behavioral Scientist* 57: 966–982.

González-Bailón, S. and Wang, N. 2016. Networked discontent: the anatomy of protest campaigns in social media. *Social Networks* 44: 95–104.

Gruzd, A., Wellman, B. and Takhteyev, Y. 2011. Imagining Twitter as an imagined community. *American Behavioral Scientist* 55: 1294–1318.

Hagberg, A., Swart, P. and S Chult, D. 2008. *Exploring Network Structure, dynamics, and Function using NetworkX(No. LA-UR-08-05495; LA-UR-08-5495)* Los Alamos National Lab. (LANL), Los Alamos, NM.

Hamzehei, A., Jiang, S., Koutra, D., Wong, R. and Chen, F. 2017. Topic-based social influence measurement for social networks. *Australasian Journal of Information Systems* 21: 61.

Harris, J. K., Duncan, A., Men, V., Shevick, N., Krauss, M. J. and Cavazos-Rehg, P. A. 2018. Peer Reviewed: Messengers and messages for tweets that Used #thinspo and #fitspo Hashtags in 2016. *Preventing Chronic Disease* 15, e01, doi: 10.5888/pcd15.170309.

Harris, J. K., Moreland-Russell, S., Choucair, B., Mansour, R., Staub, M. and Simmons, K. 2014. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. *Journal of Medical Internet Research* 16: e238.

Hilton, S. and Hunt, K. 2011. UK newspapers' representations of the 2009–10 outbreak of swine flu: one health scare not over-hyped by the media?. *Journal of Epidemiology and Community Health* 65: 941–946.

Himelboim, I., Lariscy, R. W., Tinkham, S. F. and Sweetser, K. D. 2012. Social media and online political communication: the role of interpersonal informational trust and openness. *Journal of Broadcasting & Electronic Media* 56: 92–115.

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B. and Espina, C. 2017. Classifying Twitter topic-networks using social network analysis. *Social Media+ Society* 3.

Hinds, P. and McGrath, C. 2006. Structures that work: social structure, work structure and coordination ease in geographically distributed teams. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work,* pp. 343–352.

Homeland Security Council (US) 2006. *National Strategy for Pandemic Influenza: Implementation Plan*. Executive Office of the President.

Hossain, L. and Kuti, M. 2010. Disaster response preparedness coordination through social networks. *Disasters* 34: 755–786.

Kaalia, R. and Rajapakse, J. C. 2019. Refining modules to determine functionally significant clusters in molecular networks. *BMC Genomics* 20: 901.

Kaur, M. and Singh, S. 2016. Analyzing negative ties in social networks: a survey. *Egyptian Informatics Journal* 17: 21–43.

Kearney, M. W. 2019. rtweet: collecting and analyzing Twitter data. *Journal of Open Source Software* 4: 1829.

Kemp, S. 2020. Digital 2020: April Global Statshot, available at: https://datareportal.com/reports/digital-2020-april-global-statshot (accessed May 24, 2020).

Keshvari, M., Yamani, N., Adibi, P. and Shahnazi, H. 2018. Health journalism: health reporting status and challenges. *Iranian Journal of Nursing and Midwifery Research* 23: 14.

Kruikemeier, S. 2014. How political candidates use Twitter and the impact on votes. *Computers in Human Behavior* 34: 131–139.

Liang, H., Fung, I. C. H., Tse, Z. T. H., Yin, J., Chan, C. H., Pechta, L. E. and Fu, K. W. 2019. How did Ebola information spread on twitter: broadcasting or viral spreading?. *BMC Public Health* 19: 438.

Littau, J. and Jahng, M. R. 2016. Interactivity, social presence, and journalistic use of Twitter. # *ISOJ Journal* 6: 71–90.

Lossio-Ventura, J. A. and Alatrista-Salas, H. (Eds), 2017. *Information Management and Big Data: Second Annual International Symposium, SIMBig 2015, Cusco, Peru, September 2-4, 2015, and Third Annual International Symposium, SIMBig 2016, Cusco, Peru, September 1-3, 2016, Revised Selected Papers* (Vol. 656), Springer.

Majmundar, A., Allem, J. P., Cruz, T. B. and Unger, J. B. 2018. The why we retweet scale. *PLoS ONE* 13: e0206076, available at: https://doi.org/10.1371/journal.pone.0206076.

Martin, J. G. III 2012. *Visualizing the Invisible: Application of Knowledge Domain Visualization to the Longstanding Problem of Disciplinary and Professional Conceptualization in Emergency and Disaster Management.* Universal-Publishers, Charles Town, MA.

Matsa, K. E. and Shearer, E. 2018. News use across social media platforms 2018. *Pew Research Center*.

Moody, J., McFarland, D. and Bender-deMoll, S. 2005. Dynamic network visualization. *American Journal of Sociology* 110: 1206–1241.

Morris, M. R., Counts, S., Roseway, A., Hoff, A. and Schwarz, J. 2012. Tweeting is believing? Understanding microblog credibility perceptions. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pp. 441–450.

Recuero, R., Zago, G. and Soares, F. 2019. Using social network analysis and social capital to identify user roles on polarized political conversations on Twitter. *Social Media+ Society* 5: 205630511984874.

Shin, J., Jian, L., Driscoll, K. and Bar, F. 2017. Political rumoring on Twitter during the 2012 US presidential election: rumor diffusion and correction. *New Media & Society* 19: 1214–1235.

Shin, J., Jian, L., Driscoll, K. and Bar, F. 2018. The diffusion of misinformation on social media: temporal pattern, message, and source. *Computers in Human Behavior* 83: 278–287.

Staniland, K. and Smith, G. 2013. Flu frames. *Sociology of Health & Illness* 35: 309–324.

Takeichi, Y., Sasahara, K., Suzuki, R. and Arita, T. 2015. Concurrent bursty behavior of social sensors in sporting events. *PLoS ONE* 10: e0144646.

Valente, T. W. 1995. *Network Models of the Diffusion of Innovations*, Hampton Press, Cresskill, NJ.

Valente, T. W. 2010. *Social Networks and Health: Models, Methods, and Applications,* Oxford University Press.

Wang, J., Cellary, W., Wang, D., Wang, H., Chen, S. C., Li, T. and Zhang, Y. (Eds), 2015. *Web Information Systems Engineering–WISE 2015*: *16th International Conference*, *Miami, FL, USA, November 1–3, 2015, Proceedings* (Vol. 9418), Springer.

Wojcik, S. and Hughes, A. 2019. *Sizing up Twitter users* Pew Research Center, Washington, DC.

World Health Organization. 2009. *Pandemic Influenza Preparedness and Response: A WHO Guidance Document,* World Health Organization, Geneva.

Yang, J. and Counts, S. 2010. Predicting the speed, scale, and range of information diffusion in Twitter. *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 10: 355–358.

Zhou, L., Zhang, D., Yang, C. C. and Wang, Y. 2018. Harnessing social media for health information management. *Electronic Commerce Research and Applications* 27: 139–151.