Computers in Human Behavior 89 (2018) 111–120

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Diffusion size and structural virality: The effects of message and network features on spreading health information on twitter

Jingbo Meng[a,*], Wei Peng[b], Pang-Ning Tan[c], Wuyu Liu[d], Ying Cheng[a], Arram Bae[a]

[a] Department of Communication, Michigan State University, East Lansing, USA
[b] Department of Media and Information, Michigan State University, East Lansing, USA
[c] Department of Computer Science and Engineering, Michigan State University, East Lansing, USA
[d] Department of Communication, Bellarmine University, Louisville, KY, USA

ARTICLE INFO

Keywords:
Information diffusion
Structural virality
Social media
Health information
Social network

ABSTRACT

Relying on diffusion of innovation theory, this study examines the impacts of perceived message features and network characteristics on size (i.e., the number of retweets a message receives) and structural virality (i.e., quantified distinction between broadcast and viral diffusion) of information diffusion on Twitter. The study collected 425 unique tweets posted by CDC during a 17-week period and constructed a diffusion tree for each unique tweet. Findings indicated that, with respect to message features, perceived efficacy after reading a tweet positively predicted diffusion size of the tweet, whereas perceived susceptibility to a health condition after reading a tweet positively predicted structural virality of the tweet. Perceived negative emotion positively predicted both size and structural virality. With respect to network features, the level of involvement of brokers in diffusing a tweet increased the tweet's structural virality. Theoretical and practical implications were discussed on disseminating health information via broadcasting and viral diffusion on social media.

1. Introduction

Information diffusion is a process through which information spreads through communication channels from a person or an organization to another within a social system over time (Rogers, 2003). Sharing information with one's social network has become the fundamental and constitutive activity on social media (John, 2012), which makes social media a desirable space for spreading information. The value of social media in information diffusion lies not only in its ability to broadcast information to a large number of people but also in its support for social networks through which information can travel and reach more people who are otherwise not exposed to the information. Information diffused the most can signal the importance of the information at a particular time and quickly focus the public's attention on the issue (Nahon & Hemsley, 2013). In the context of health communication, health organizations have turned to social media to diffuse health education and prevention information (Harris, Mueller, & Snider, 2013). Identifying factors that drive the diffusion of health information is critical for scholars to understand how health information may or may not spread in a similar way as other viral content on social media, as well as for practitioners to design spreadable health messages that need the public's attention and action.

Communication scholars have focused on identifying message features (e.g., information utility) that drive information diffusion in social media (Berger & Milkman, 2012; Kim, 2015; McLaughlin, Hou & Meng, 2016). However, classic diffusion research acknowledges the importance of interpersonal networks in the process of information diffusion (Katz & Lazarsfeld, 1955), which has been supported by information science scholars (Bakshy, Rosenn, Marlow, & Adamic, 2012; Watts & Dodds, 2007). Diffusion of Innovations (DOI) theory argues that diffusion is determined by not only attributes of an innovation (e.g., novelty and relative advantage) but also properties of communication networks (e.g., opinion leaders and weak ties) within which the innovation spreads (Rogers, 2003). On social media such as Twitter, health information could present different message features and be passed along by users with different network positions or roles in the communication network (Figueiredo, Chen, & Azevedo, 2015). Therefore, in order to have a complete understanding of information sharing in the social media environment, this study examines the effects of both message and network features on diffusion of health-related information.

More importantly, this study aims to advance the research on online information diffusion by examining both the size and structure of diffusion as outcomes. Diffusion size is the aggregated number of

* Corresponding author. 404 Wilson Road, RM 481, Communication Arts and Sciences Building, Michigan State University, East Lansing, MI 48824, USA.
E-mail address: jingbome@msu.edu (J. Meng).

https://doi.org/10.1016/j.chb.2018.07.039
Received 27 November 2017; Received in revised form 16 June 2018; Accepted 27 July 2018
Available online 28 July 2018

adoptions of an innovation over time (Rogers, 2003), while diffusion structure characterizes patterns of diffusion of an innovation (Goel, Anderson, Hofman, & Watts, 2016). Diffusion structure is an important dimension for understanding a cascade on a continuum with broadcast and viral spreading as the two extreme patterns (Goel et al., 2016). The examination of structural virality presents an extension of DOI theory beyond the size of diffusion. In this study, the innovations examined are new tweets originally posted by Centers for Disease Control and Prevention (CDC) on Twitter. Adoption of the innovation is retweeting behavior of an individual user. This study aims to investigate message and network features that affect the size and structure of diffusion of health information tweeted by CDC.

2. Literature review

2.1. Diffusion of innovations and structural virality

DOI theory explains the process by which an innovation propagates in a social system over time (Rogers, 2003). It theorizes attributes of innovations, diffusion networks and adopter categories that affect the adoption of an innovation. DOI theory has been applied to research on diffusion of health information on social media, such as drug news on Twitter (McLaughlin et al., 2016) and health intervention messages on Facebook (Kee, Sparks, Struppa, Mannucci, & Damiano, 2016). On social media, sharing information via personal networks is a constitutive activity motivated by users' gratifications (John, 2012). In this competitive and saturated information environment, social sharing behavior such as retweeting or passing along information indicates a person's approval or acceptance of the piece of information (Kee et al., 2016).

Diffusion studies built on DOI theory typically focus on the aggregated number of adoptions (i.e., size) as a diffusion outcome (Valente, 1995). Relatively little attention has been paid to explain structural patterns of diffusion. The accumulative number of adoptions may arise from two distinct structural mechanisms: broadcast and viral diffusion (Goel et al., 2016). This idea of broadcast versus viral spreading is not new. Back to Roger's theoretical discussion on communication channels for diffusion and the classic two-step flow model (Katz & Lazarsfeld, 1955), broadcast via mass media and virality via interpersonal connections are the two primary routes for information to diffuse in a community. Structural virality formally conceptualizes structural patterns of diffusion by quantifying the distinction between broadcast and viral diffusion (Goel et al., 2016). Broadcast diffusion depicts a pattern where a large number of adoptions grow from a single parent node, whereas viral diffusion depicts a pattern where multiple generations contribute to the diffusion process and any one node infects only a few others (Goel et al., 2016). Diffusions with the same size may present structural diversity characterized by both broadcast and viral mechanisms (Fig. 1), as well as conceivable combinations of the two. An analysis with 1 billion unique tweets showed that a higher correlation between size and structural virality was only 0.2 (Goel et al., 2016). Therefore, an examination of structural virality contributes to our theoretical and empirical understanding of fine-grained structures of diffusion as an outcome variable beyond the aggregated size of diffusion. Fig. 1 presents a visual example with different patterns of diffusion.

2.2. Perceived message features and information diffusion

2.2.1. Information utility: threat and efficacy

DOI theory claims that perceived attributes of an innovation are an important explanation of diffusion of innovations (Rogers, 2003). Following this logic, perceived message features could determine the extent to which individuals share the messages to their social circles. Unlike expressed message features that focus on literal message variations that are independent of audience perceptions or responses, *perceived* message features are effect-based in terms of audience responses towards messages (O'Keefe, 2003). Relying on DOI theory, the current study focuses on the effect of perceived, rather than expressed, message features on diffusing health-related tweets. Research has shown that perceived information utility of a message is positively associated with individuals' sharing behavior (Kim, 2015). However, it is not clear what message features constitute perceived information utility. Information utility model (Knobloch, Carpentier, & Zillmann, 2003) helps to unpack this audience perception. The model claims that the susceptibility and the severity of experiencing negative events suggested in a message indicate its information utility. For example, a tweet message “The rate of new cases of melanoma is expected to double by 2025” may make people perceive highly susceptible to melanoma; a tweet message “Melanoma is the deadliest form of skin cancer, killing 9000 people each year” may make people perceive high severity of having melanoma. Individuals need to monitor their environment to be able to adapt and to survive, and therefore, information that make people perceive threats is considered to be more useful. On social media, impression management and self-presentation shape people's communication behavior (Boyd & Ellison, 2007). Passing on useful information enhances an individual's image as being well informed, smart and helpful (Berger, 2014). Communicating that desired image motivates individuals to spread information of high utility to their social contacts.

In addition to impression management based on information utility, persuading others is another psychological mechanism that could explain information sharing with others (Berger, 2014). Research has shown that people have concerns for others in their social networks and want to help save others from negative experience (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004). Passing on information is a way of exerting interpersonal influence. In persuasion literature, the extended

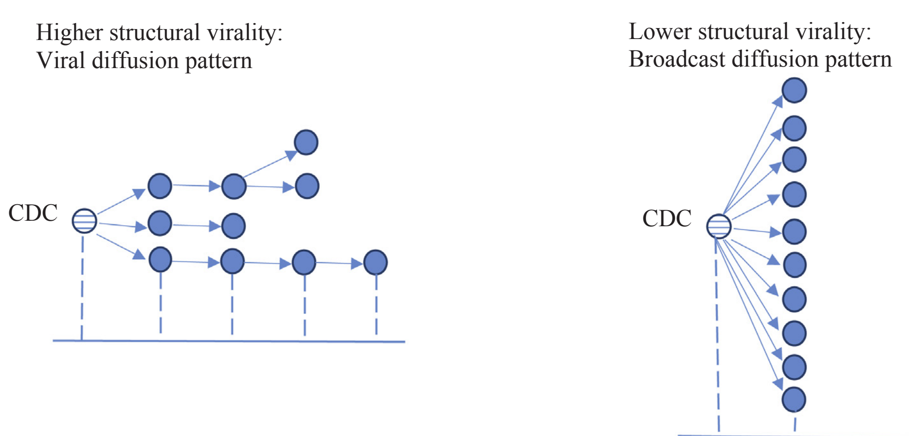

Fig. 1. A Diffusion with the same size but different structural virality. Note. Each solid circle is a retweeter.

parallel process model (Witte, 1992) argues that messages that produce high levels of perceived susceptibility and severity of experiencing a negative event is more persuasive than messages characterized as low in perceived susceptibility and severity. The above arguments from two different theoretical perspectives both indicate that tweet messages of high information utility and of high persuasiveness (i.e., information conveying susceptibility and severity) should be more likely to be retransmitted.

Hypothesis 1 (H1). Health-related tweets that make individuals perceive higher (a) susceptibility and (b) severity of experiencing negative health conditions have a larger size of diffusion.

According to the information utility model (Knobloch et al., 2003), efficacy information often provides content that is of high information value. Efficacy typically includes response efficacy and self-efficacy. Response efficacy refers to the effectiveness of the recommended response in deterring health risks (Witte et al., 1996). Self-efficacy refers to one's capability of performing the recommended response (Witte et al., 1996). Response efficacy information may address benefits of performing a recommended solution, while self-efficacy information may explain simple steps about how to carry out the recommended solution to minimize health risks. For example, a tweet message "2 doses of MMR vaccine are 97% effective" may make people perceive high response efficacy of MMR vaccine; a tweet message "You can protect yourself from Hepatitis A when travelling with a simple #vaccine …" may make people perceive high self-efficacy to take the vaccine. Past studies have shown that messages with high-level of efficacy information were rated as more useful (Knobloch-Westerwick & Sarge, 2015). News stories that increased efficacious perceptions were retransmitted more times via emails (Kim, 2015). Therefore, we hypothesize that:

Hypothesis 2 (H2). Health-related tweets that make individuals perceive higher (a) response efficacy and (b) self-efficacy in dealing with health conditions have a larger size of diffusion.

Although previous literature is informative in explaining the aggregated size of information diffusion, it is not clear how message features based on information utility influence the structural patterns of diffusion. Therefore, we raise the following questions:

Research Question 1 (RQ1): Do message features based on information utility, including perceived (a) susceptibility, (b) severity, (c) response efficacy and (d) self-efficacy, influence structural virality?

2.2.2. Emotion: positivity and negativity

Beyond information utility, emotionally charged content has been documented to be more viral on social media (Stieglitz & Dang-Xuan, 2013). Studies on emotional valence and information diffusion have yielded mixed findings. A couple of studies reported that emotional positivity, operationalized as the difference between the percentage of positive and negative words in a new article, predicted the amount of news transmission via email (Berger & Milkman, 2012; Kim, 2015). Others found that health-related tweets with positive or negative affective tone were more likely to be propagated than the ones with neutral affective tone (McLaughlin et al., 2016), and that spikes of tweet volume about an event were associated with increased negative sentiment contained in tweets (Thelwall et al., 2011). While positive messages are argued to be passed on more frequently because they make recipients upbeat and enhance the sharers' positive images (Kim, 2015), the negativity bias is also well documented in the theory of news values (Galtung & Ruge, 1965; Vosoughi, Roy, & Aral, 2018).

Emotional messages may elicit greater cognitive involvement such as attention (Kissler et al., 2007), which in turn, leads to a higher likelihood of behavioral response in the form of information sharing (Rimé, 2009). Moreover, social sharing is a fruitful way to regulate one's emotion (Gross & John, 2003). Sharing positive emotions reconsumes the positive affect and extends it to others. People in general like to be around positive persons, and thus, sharing positive emotions helps build social bonding (Berger, 2014). On the other hand, sharing negative emotion is beneficial in that it improves one's mood, such as reducing anxiety or feeling of dissonance (Berger, 2014). People may feel better and receive social support after expressing negative emotion (Grandey, 2000). Communicating negative emotion can also be construed as showing out of the ordinary or cynical sophistication (Cappella, Kim, & Albarracin, 2015). Therefore, we hypothesize that:

Hypothesis 3 (H3). Health-related tweets that invoke higher levels of (a) positive emotion and (b) negative emotion have a larger size of diffusion.

Similarly, even though literature has discussed the effects of emotional content on virality (Berger & Milkman, 2012; Stieglitz & Dang-Xuan, 2013), virality has been generally defined as the size of diffusion rather than structural virality. Therefore, we raise the following questions:

Research Question 2 (RQ2): Do message features based on emotion, including perceived (a) positive emotion and (b) negative emotion, influence structural virality?

2.3. Network features and information diffusion

2.3.1. Opinion leaders

In addition to perceived innovation attributes, DOI theory posits the importance of interpersonal diffusion network in the spread of an innovation (Rogers, 2003). Opinion leaders are the few individuals who have the largest number of social ties in a diffusion network (Valente, 1995). People are inclined to monitor information sent by opinion leaders as well as emulate their behavior as opinion leaders are in an advantaged position for access to information and scanning the environment in the community (Valente, 1995). In other words, opinion leaders represent the norm of the community and are respected as influential. According to the two-step flow model of communication (Katz & Lazarsfeld, 1955), opinion leaders act as intermediaries between mass media and the majority of people in a community. Therefore, the mechanism that opinion leaders facilitate information diffusion is that they stimulate adoption behavior (e.g., retweet a message) of their followers. Followers' information sharing help create multiple generations of adopters and extend the diffusion chain. Previous studies have showed that opinion leaders were more important than average individuals in diffusing public opinions (Watts & Dodds, 2007), and opinion leaders were critical in accelerating behavioral diffusion (Valente, 1995). Thus, we speculate that.

Hypothesis 4 (H4). For the diffusion of each unique tweet, the level of involvement of opinion leaders positively predicts its (a) size and (b) structural virality.

2.3.2. Brokers

Recent development of DOI theory argues for the important role of brokers in creating diffusion cascades (González-Bailón, Borge-Holthoefer, & Moreno, 2013). A critical property of online networks is that actors tend to connect with those already connected to their neighbors, leading to local clustering than one would expect by chance (Newman, 2010). Brokers are individuals who bridge distinct clusters or sub-communities in a social network (Burt, 1992). Individuals with high levels of brokerage can reduce the overall distance between others in a network, increasing the likelihood and efficiency of information diffusion in a network (Valente & Fujimoto, 2010). Moreover, bridging ties tend to be weak ties (Easley & Kleinberg, 2010), characterized by the low frequency of contacts and low emotional intensity involved in the relationship (Granovetter, 1973). Weak ties are argued to facilitate information diffusion in that they transmit new and non-redundant information, which is more likely to be picked up by loosely connected groups of individuals (Granovetter, 1973).

Extensive literature has supported the critical role of brokers in transmitting information to a larger scale (Yang & Counts, 2010). For example, a recent study reported that the top 1% of brokers in the follower-followee network controlled 25% of all the information transmission on Twitter (Lou & Tang, 2013). Brokers could show an even greater influence on structural virality given their positional characteristics bridging otherwise unconnected groups. Due to the information asymmetry and low-redundancy between unconnected groups, brokers are able to pass and infuse new information to distinct groups, which helps extend the chain of information diffusion. Research has shown that top brokers controlled almost 80% of the information transmission between different clusters (Lou & Tang, 2013), which primarily determined how far information could travel from its source (Yang & Counts, 2010). Moreover, brokers tend to have more weak ties, and weak ties retransmit information significantly more than strong ties (Bakshy et al., 2012). In other words, information passed along by brokers has a higher chance of being retransmitted in the network, which further extends the diffusion chain (González-Bailón et al., 2013; Zhao et al., 2012). Therefore, we propose that.

Hypothesis 5 (H5). For diffusion of each unique tweet, the level of involvement of brokers positively predicts its (a) size and (b) structural virality.

3. Methods

3.1. Overview

This study collected original tweet messages posted by the official account of CDC on Twitter from 7 April 2015 to 4 August 2015 (17 weeks) and the Twitter users who have retweeted those messages. Twitter search API (https://dev.twitter.com/rest/public/search) was used to fetch CDC's original tweets and retweet data. Starting on 7 April 2015, the API was used to download each new tweet posted by CDC and searched the exact tweet message to track the number of retweets and retweeters of the new tweets for 14 days[1]. This process resulted in 425 unique tweets and 10,035 retweeters.

The unit of analysis is each unique tweet message. To evaluate perceived message features, we conducted an online survey to crowdsource individual perceptions after reading tweet messages. To compute network features, we collected each retweeter's number of followers on Twitter and the follower-followee network of the 10,035 retweeters. To compute diffusion outcomes, we collected the number of retweets received by each unique tweet and constructed a diffusion tree for each unique tweet for calculating structural virality.

3.2. Data collection and measures for perceived message features

Perceived message features were measured by an online message evaluation survey where respondents read and rated the 425 CDC twitter messages (O'Keefe, 2003). The goal of this online survey was to crowdsource evaluations of perceived information utility and emotion for each tweet by aggregating ratings of multiple respondents who read the same tweets (Kim, 2015). The online survey was administered through Amazon's Mechanic Turk. Respondents recruited through MTurk are often more representative of the U.S. population than convenience samples (for details about the validity of using MTurk samples, see Berinsky, Huber, & Lenz, 2011). A total of 3310 U.S. adults completed the online survey. Of the respondents, 49.8% were female, and the average age was 33.5 (*SD* = 10.33). The majority of the participants were Caucasians (78.3%). Three out of the 425 Twitter messages were randomly displayed to each participant for evaluation using an online survey service Qualtrics. On average, each tweet received evaluations from 23.23 (*SD* = 1.08) independent respondents.

Each Twitter message was evaluated by three sets of questions. The first set of questions adapted items from the Risk Behavior Diagnostic Scale (Witte et al., 1996) to measure participants' perceived susceptibility, severity, response efficacy, and self-efficacy after reading a tweet message. Three items were used to measure each variable on a 5-point scale ranging from strongly disagree (=1) to strongly agree (=5) with item statements that described their perceptions. For example, after presenting a tweet message, "Don't eat, serve, or sell any Blue Bell products. New info on Listeria outbreak …" participants were instructed, "After reading the above tweet, please indicate to what extent you agree or disagree with the following statements." An example item for susceptibility was, "I am likely to get impacted by the health issue mentioned in this tweet." An example item of severity was, "The health issue mentioned in this tweet is a serious problem." An example item of response efficacy was, "This tweet suggests an effective way to solve the mentioned health issue." An example of self-efficacy was, "I feel I can do what is recommended in the tweet." The measures yielded a good internal reliability for susceptibility ($\alpha = 0.95$), severity ($\alpha = 0.88$), response efficacy ($\alpha = 0.95$) and self-efficacy ($\alpha = 0.96$).

The second set of questions measured perceived positive and negative emotions. It asked participants to rate on a series of single items about specific emotions felt while reading a tweet message (including the image if there was any). Participants were asked to rate on a 5-point scale ranging from not at all (= 1) to extremely (= 5) about positive emotions, including joy, love, hope, relief, compassion, and negative emotions, including fear, sadness, disgusting and anger (Lazarus, 1991). The third set of questions asked participants to rate how novel and entertaining a tweet message was. One-way random effects ICC(1, k) = 0.67 with 95% confidence interval [0.63, 0.70], indicating moderate to good inter-rater reliability (McGraw & Wong, 1996; Shrout & Fleiss, 1979).

A principal component analysis with promax rotation was conducted with items that assessed positive and negative emotions. It generated two factors that were consistent with the conceptual categorization of positive and negative emotions (Lazarus, 1991). Joy, love, hope, relief and compassion had factor loadings at 0.83, 0.89, 0.92, 0.78, and 0.90 respectively on positive emotion; fear, sadness, disgust and anger had factor loadings at 0.83, 0.85, 0.76 and 0.90 respectively on negative emotion. Then, two composite indices were computed by averaging items for positive ($\alpha = 0.90$) and negative emotions ($\alpha = 0.84$).

3.3. Data collection and measures for network features

The Twitter search API was also used to fetch CDC's followers, retweeters' followers and followees who also retweeted any of the 425 unique tweets to construct a network involving follower-followee relationships among all the retweeters. Data on the follower-followee relationships were collected every 14 days after all the retweeters of a unique tweet were collected.

To measure the level of involvement of brokers, we first identified the follower-followee ties among all the retweeters with 217,624 ties in one large component. Then, Louvain method was used to detect clusters in the network. Louvain method was selected given its high efficiency and capacity to decompose a large network into mutually exclusive clusters that seek to maximize modularity (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008, p. P1008). Lastly, the number of clusters to which a retweeter's neighbors belong indicates the degree to which the retweeter spans local clusters (Fleming & Waguespack, 2007). For each unique tweet, the level of involvement of brokers in the diffusion process is the average number of clusters spanned by its retweeters.

In-degree centrality, defined as the number of incoming ties, is the most frequently used network measure to identify opinion leaders in a network (Valente, 1995). Individuals with high in-degree centrality in the Twitter network may be considered as prestigious or expert in providing information of great value. Therefore, the number of followers on Twitter was used as a measure of opinion leader. For each unique tweet, the level of involvement of opinion leaders in the

Table 1
Examples of tweets, their perceived message feature ratings, and diffusion outcomes.

ID	Tweet message	Message feature ratings						Diffusion outcomes	
		Susceptibility	Severity	Response efficacy	Self-efficacy	Positive emotion	Negative emotion	Number of retweets	Structural virality
1	New #outbreak: Salmonella infections from raw frozen stuffed chicken entrees: http://t.co/jPVbi90IZt	3.26	3.64	1.21	1.33	1.04	2.89	110	2.97
2	Melanoma is the deadliest form of skin cancer, killing 9000 people each year. #VitalSigns http://t.co/rm7FNxlmBA	3.00	4.31	2.79	3.02	1.23	2.02	97	2.07
3	2.5 h of physical activity per week has health benefits. Stay healthy & safe while swimming! #SwimHealthy http://t.co/usa02e9kv1	2.06	3.00	4.34	4.10	2.54	1.03	60	1.91
4	When your child is sick, don't leave #medicine by their bed for the next dose. Keep #medsupaway & out of sight #NSM15 http://t.co/lEcoo2PAcS	2.01	3.61	3.94	4.52	2.02	1.61	75	2.06
5	53% of U.S. kids who died from heatstroke were forgotten in cars. Act fast. Save a life. #heatstrokekills http://t.co/MtBooWhKFj	1.82	3.96	2.86	3.82	2.07	3.21	88	2.48
6	Community cancer prevention saves lives & could save $2.7B in treatment by 2030. http://t.co/rm7FNxlmBA #VitalSigns	2.98	4.20	3.42	3.32	2.76	1.55	29	1.00
7	Don't eat, serve, or sell any Blue Bell products. New info on Listeria outbreak: http://t.co/cV3ogQcKFA http://t.co/d6g0MaJ0eT	3.01	3.79	3.63	4.21	1.31	3.08	473	3.61
8	Polio is still a threat in some countries. Protect kids w/polio vaccine including before international travel. http://t.co/8eiVNyhmG8	2.07	3.75	3.81	4.14	2.05	2.02	90	1.13
9	How antibiotic resistant germs spread from farm to the table http://t.co/P39QeeNPts http://t.co/azxJjQDtne	3.34	3.71	1.79	1.84	1.35	2.65	329	3.81
10	Four tips to protect against food poisoning when eating out: http://t.co/pA621ZwIry #SafeFood http://t.co/abOEuDcsHK	3.81	3.93	3.35	3.39	1.54	2.93	93	5.81
11	CDC's Dr. Anne Schuchat & @CDC_TB will chat w. @Dr.RichardBesser on Tue, 1 p.m. ET, discussing XDR #TB and #MERS. Join us at #abcDrBchat	1.90	2.90	2.21	2.37	1.31	1.16	13	1.18
12	Meet CDC #DiseaseDetective Jeff who traveled to 100 + health centers & hospitals in Sierra Leone to fight #Ebola. http://t.co/YuGLxcJJdj	1.67	3.81	2.57	2.33	2.75	1.24	33	1.67

Note. Each perceived message feature was measured on a 5-point scale. The ratings presented were the average scores. Higher scores indicated higher levels of perceived message features.

Table 2
Zero-order correlation among key variables.

		1	2	3	4	5	6	7	8	9	10	11	12
1	Susceptibility	1											
2	Severity	**0.43**	1										
3	Efficacy	**0.34**	**0.18**	1									
4	Positive emotion	**−0.12**	**0.19**	0.03	1								
5	Negative emotion	**0.18**	**0.45**	−0.08	**−0.17**	1							
6	Opinion leader	−0.06	−0.03	−0.08	0.02	0.02	1						
7	Broker	**−0.13**	−0.02	−0.07	**0.13**	−0.07	**0.26**	1					
8	Novelty	**0.11**	−0.07	**0.31**	0.08	**0.21**	0.05	**0.18**	1				
9	Entertaining	0.09	**−0.20**	0.01	**0.19**	**−0.24**	0.02	−0.07	**0.10**	1			
10	Visualization	−0.02	**0.18**	0.05	**0.21**	**0.12**	−0.06	−0.05	**0.11**	**0.22**	1		
11	Size	**0.19**	0.07	**0.17**	**−0.12**	**0.21**	−0.03	−0.03	0.03	0.06	**0.18**	1	
12	Structural Virality	**0.19**	0.07	0.08	−0.06	**0.12**	−0.02	***0.38***	−0.10	0.05	**0.11**	**0.36**	1
	Mean	2.23	3.28	3.23	1.86	1.49	896.95	3.58	2.92	2.28	0.53	39.89	2.26
	SD	0.42	0.58	0.59	0.39	0.34	799.02	0.55	0.44	0.41	0.49	42.98	0.63

diffusion process is the average number of followers of its retweeters.

3.4. Measures for diffusion outcomes

To measure the size of information diffusion, we used the number of retweets for a unique tweet in 14 days after it was first posted. Given that 94.97% of all retweets happen within the first day of the original tweet being published (Sysomos, 2010), a 14-day tracking gives us confidence in capturing the size of diffusion.

To compute structural virality, we first constructed a diffusion tree for each unique tweet following the method described in Goel et al. (2016). Specifically, for each unique tweet whose diffusion we seek to trace, we recorded (1) the retweeter (i.e., the identity of the user who retweeted the tweet); (2) the retweet time (i.e., the time at which the adoption happened); and (3) the identities of all users the retweeter follows – hereafter referred to as the retweeter's followees – from whom the retweeter could conceivable have learned about the tweet. For each unique tweet (i.e., each diffusion event), the followee who tweeted the tweet most recently before the focal retweeter was labeled the focal retweeter's "parent". In our dataset, this method allowed us to trace back to the original author of each unique tweet – that is CDC – for all 425 tweets. We then computed structural virality v(T) as the average distance between all pairs of nodes in a diffusion tree T (Goel et al., 2016); that is, for $n > 1$ nodes,

$$v(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

where d_{ij} denotes the length of the shortest path between nodes i and j. The metric v(T) provides a continuous measure of structural virality, with higher values indicating viral diffusion whereas lower values indicating broadcast diffusion (see more details in Appendix A).

3.5. Control variables

Previous research has indicated that information diffusion can be contingent upon other message features including novelty (Kim, 2015), entertainment, visualization, word count, health topics, inclusion of hashtags and URLs in tweets (Suh et al., 2010). Therefore, these extra message features were controlled in the statistical analysis. To measure these controlled message features, we asked survey respondents to rate to what extent the information in a tweet was novel and entertaining on a 5-point scale (1 = not at all, 5 = extremely). The two trained coders (see Appendix for more details) coded the presence of an image underneath a tweet, the presence of hashtags and URLs in a tweet. The two coder reached an agreement of 99% in coding these content features. In addition, the two coders were asked to write down notes about names of diseases mentioned in each tweet and primary purposes of the tweets (e.g., education, prevention, announcement, etc.) Based on these brief notes, the two coders discussed and inductively generated two broad categories of health topics: health education and prevention, and CDC announcements of events.

3.6. Data analysis

Linear regression models were used to test hypotheses and research questions with structural virality as the dependent variable. The other dependent variable, the size of diffusion was a count variable, reflecting the number of occurrences of a behavior in a period of time. Count dependent variables can violate the assumptions of linear regression, such that they often display heteroscedasticity and non-normal conditional distributions of errors (Cameron & Trivedi, 2013). Therefore, negative binomial regression was used to overcome the non-normal distribution of the number of retweets.

4. Results

4.1. Descriptive results

Among the 425 original tweets posted by CDC, the size of information diffusion (i.e., number of retweets) ranged from 1 to 473 ($M = 39.89$, $SD = 42.98$, skewness = 6.89, kurtosis = 38.73) and structural virality ranged from 1.00 to 5.81 ($M = 2.26$, $SD = 0.63$, skewness = 1.08, kurtosis = 5.58). Consistent to findings from a dataset of 1 billion tweet events (Goel et al., 2016), median structural virality was less than 3, exhibiting fairly shallow diffusion trees. For the perceived message features of the 425 tweets, the average ratings were 2.22 ($SD = 0.42$) for susceptibility, 3.28 ($SD = 0.58$) for severity, 3.23 ($SD = 0.59$) for efficacy[2], 1.86 ($SD = 0.39$) for positive emotion, 1.49 ($SD = 0.34$) for negative emotion, 2.92 ($SD = 0.44$) for novelty, and 2.28 ($SD = 0.41$) for entertaining. Among all the tweets, 82.9% included URLs and 85.6% included hashtags. Table 1 presents message examples selected based on the survey ratings. Table 2 presents descriptive statistics of and zero-order correlations among the key variables in this study.

With respect to health topics, 79.92% of the tweets communicated health education and prevention knowledge; 20.08% of the tweets featured CDC's announcement such as sponsored events and awards. Among the tweets that mentioned health education and prevention knowledge, specific topics covered included cancer (39), vaccine (31), food safety (29), ebola (27), drug overdose (24), swim safety (19), MERS (Middle East respiratory syndrome, 19), tick/lyme disease (17), HIV (15), mosquito-borne disease (11), and others (e.g., diabetes, birth control, hepatitis, heart disease, hand hygiene, injury prevention, and listeria outbreak etc.)

4.2. Impacts of message and network features on diffusion size

H1 posited positive effects of perceived susceptibility and severity of experiencing a negative health condition as message features on the size of diffusion. The result showed that neither perceived susceptibility ($p = .09$) nor severity ($p = 0.14$) had a significant effect on the number of retweets that a unique retweet received. Therefore, H1 was not supported. H2 stated a positive effect of perceived efficacy as a message feature on the size of diffusion. The result showed that perceived efficacy after reading a tweet message positively predicted the number of retweets received by the message, $B = 0.14, p < .05$. It indicated that a message would be retweeted 1.15 ($e^{0.14}$) times more with one unit increase in the efficacy perceived by respondents. Therefore, H2 was supported. H3 posited effects of positive and negative emotions on the size of diffusion. The result showed that negative emotion as a message feature positively predicted the number of retweets received by the message, $B = 0.22, p < .01$, whereas positive emotion was not a significant predictor, $p = .29$. This indicated that a message would be retweeted 1.25 ($e^{0.22}$) times more with one unit increase in perceived negative emotion. Therefore, H3b was supported while H3a was not.

H4a and H5a posited that the level of involvement of opinion leaders and brokers in the diffusion process would predict the size of diffusion. The results showed that neither the level of involvement of opinion leaders ($p = 0.30$) nor brokers ($p = 0.10$) in diffusing a unique tweet had a significant effect on the number of retweets received by the message. Therefore, H4a and H5a were not supported.

4.3. Impacts of message and network features on diffusion structure

RQ1 aimed to explore the effects of message features based on perceived information utility on structural virality of diffusion. The results showed that only perceived susceptibility had a positive effect ($\beta = 0.15, p < .01$) on structural virality. Neither perceived severity ($p = .39$) nor efficacy ($p = .12$) was a significant predictor. RQ2 explored the effect of message features based on perceived positive and negative emotions on structural virality. The results showed that negative emotion had a positive effect ($\beta = 0.13, p < .05$), whereas positive emotion ($p = .29$) did not have a significant effect on structural virality. H4b and H5b posited positive effects of the level of involvement of opinion leaders and brokers on structural virality. The results showed that the level of involvement of brokers significantly increased structural virality of diffusion for a unique tweet message ($\beta = 0.45, p < .001$), whereas the level of involvement of opinion leaders did not show a significant effect ($p = .78$). Table 3 presents the results for regression models predicting diffusion size and structural virality.

5. Discussion

5.1. Theoretical implication

The present study examined the effects of message and network features on the size and structural virality of health information diffusion on Twitter. It discovered different sets of predictors in influencing the size and structural virality of diffusion respectively. Among message features based on Information Utility Theory, perceived efficacy was a positive predictor of diffusion size, whereas perceived susceptibility of experiencing a health condition was a positive predictor of structural virality. This finding is unique because it revealed that the threat component of information utility becomes critical in explaining person-to-person diffusion even after controlling for perceived negative emotion such as fear and sadness. Unlike previous literature that has emphasized efficacy as a major indicator of information utility and predictor of information transmission (Cappella, Kim, & Albarracín, 2015), our finding pointed out that for health information to go viral, perceived susceptibility of experiencing the negative event should be the core to communicate. Future research is encouraged to consider both the threat and efficacy component (Witte, 1992) of information utility when spreading other content such as environmental issues (e.g., climate change) on social media.

This study found that perceived negative emotion was, while perceived positive emotion was not, a significant predictor of size and structural virality of diffusion. While the finding is consistent with health information sharing literature on Twitter (McLaughlin et al., 2016), it contradicts sharing health news via email and Facebook (Kim, 2015). Different media platforms can affect what people talk about and share (Berger & Iyengar, 2013) and thus may explain the discrepancies in those findings. Email and Facebook networks involve a great number of pre-existing social relations (Ellison, Steinfield, & Lampe, 2011), while Twitter network may involve a larger number of strangers or very weak ties as following a Twitter user typically does not require a mutual agreement from both relational parties (Hansen, Arvidsson, Nielsen, Colleoni, & Etter, 2011). Social bonding is an important motivation when sharing information with pre-existing social relations (Berger, 2014). Sharing positive emotions helps build a positive self-image and maintain positive interpersonal relationships. However, when communicating with very weak ties, it may be less important to be a positive or a nice person. Instead, sharing negative-tone messages may attract more attentions from others. Information that arouses negative emotion may signal danger and thus are more urgent and useful (McLaughlin et al., 2016). In addition, passing along negative tweets may show a person's discriminating judgement and heightened cynical sophistication (Cappella et al., 2015), just like we tend to perceive book review writers as more intelligent and competent when they provide negative as opposed to positive reviews (Amabile, 1983). Future research is encouraged to compare diffusion of the same message on different media platforms and examine varying motivations of information sharing.

With respect to network features, neither opinion leader nor broker involvement level was a significant predictor of diffusion size. The finding does not align with DOI theory that primarily explains adoptions of products or behaviors, but is consistent with the claim of million follower fallacy, such that users with high in-degree are not influential in spawning retweets (Cha, Haddadi, Benevenuto, & Gummadi, 2010). The explanation may be that information differs from products or behaviors in that information utility lies in its non-redundancy (Granovetter, 1973), whereas multiple adoptions of the same product or behavior accelerate diffusion of that product or behavior (Centola, 2010). Users with high in-degree centrality may consider retweeting CDC's posts as spreading redundant and less useful information given that CDC already broadcasts to an enormous number of followers on Twitter. As found in Cha et al. (2010), the positive effect of in-degree centrality on spawning retweets of a user was mainly observed for Twitter users who were mainstream news organizations, but not for other users with high in-degree centrality. An alternative explanation is based on the DOI literature that opinion leaders are usually experts in specific domains (Rogers, 2003). The counts of followers may not help identify opinion leaders in health news. Or the followers of opinion leaders identified in this study may not be interested in health news, and therefore did not retransmit the news.

Moreover, this study found that the level of involvement of brokers was a positive predictor of structural virality of diffusion, indicating that the more clusters that its retweeters spanned in the diffusion network, the more likely that a unique tweet manifested a viral structure of diffusion. This finding further demonstrated the trapping effect of community structures in affecting information diffusion (Onnela et al., 2007), such that viral memes can permeate through many communities, while non-viral memes are often confined within a community (Weng, Menczer, & Ahn, 2013). Given the similarity shared among people in the same community, information tweeted by users who belong to the single community may become old and redundant very soon. In contrast, information tweeted by brokers who connect several communities may be novel and relevant to different groups of people, and thus is

Table 3
Models predicting diffusion size and structural virality.

	Model 1					Model 2			sig.
	Predicting Diffusion Size					Predicting Structural Virality			
	B	Exp(B)	Wald Chi-Square	p-value	sig.	Beta	t-value	p-value	
Perceived Message Features									
Susceptibility	0.11	1.12	2.95	0.09		0.15	2.65	0.004	**
Severity	−0.12	0.89	2.63	0.14		−0.06	−0.91	0.39	
Efficacy	0.14	1.15	4.86	0.03	*	0.05	1.04	0.12	
Positive emotion	0.06	1.06	0.92	0.29		−0.06	−1.16	0.29	
Negative emotion	0.22	1.25	8.99	0.001	**	0.13	2.81	0.01	*
Network Features									
Opinion leaders involvement level	−0.08	0.92	1.67	0.30		−0.09	−1.73	0.78	
Brokers involvement level	0.14	1.15	3.14	0.10		0.45	9.23	0.000	***
Control Variables									
Word count	0.01	1.01	0.33	0.65		0.13	0.28	0.72	
Health topic									
CDC announcement (ref.)									
Prevention/education	0.34	1.02	5.26	0.02	*	0.06	1.74	0.15	
URL	−0.16	0.85	1.32	0.34		0.17	0.38	0.69	
Hashtag	−0.14	0.87	0.77	0.15		0.02	0.50	0.72	
Visualization	0.17	1.19	9.92	0.001	**	0.11	2.25	0.02	*
Novelty	−0.06	0.94	1.03	0.34		0.09	1.63	0.14	
Entertaining	0.05	1.03	0.16	0.45		0.09	1.73	0.07	
			Pseudo-R^2: 0.29			Adjusted R^2: 0.23			

Note: (1) Multicollinearity test showed that all variance inflation factors (VIFs) were smaller than 2.15 in both models. (2) Depending on the distribution of the dependent variables, Model 1 used negative binomial regression; Model 2 used OLS regression. (3) In Model 1, Pseudo-R^2 is computed using 1 – deviance(fitted_model)/deviance(intercept_only). It indicates the proportion of deviance reduced by including current predictors compared to using no predictors. (3) Predictors were all standardized. (4) $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

more likely to be picked up and retweeted further.

5.2. Practical implication

Our findings have implications on strategies for disseminating public health information on Twitter. The insights are that to increase the aggregated number of retweets, designing efficacious information is the key; to increase the diffusion chain through person-to-person transmission, crafting information that can raise risk perception is important. Tweets that induce negative emotions could be more effective in catching users' attention and expanding sharing of the information. Practitioners may use loss frames or words that convey negative sentiments to boost the spread of the information on Twitter.

Moreover, we have found that CDC's tweets mainly diffused via broadcasting rather than person-to-person virality. Compared with viral health-related tweets, CDC's tweets are more of top-down promotional messages than bottom-up socially driven messages. For example, Madalyn Parker tweeted a message about taking sick days for mental health. The tweet went viral in a way that retweets flowed in conversing about work culture, organizational practice, and stigma surrounding the issue. To increase the virality of health-related tweets, CDC could craft messages that spark storytelling and conversations among the public, which could draw more and longer public's attention.

5.3. Limitation

This study has a few limitations. First, survey respondents recruited from MTurk may bring potential bias in the study. MTurk participants might be a subset of people who perceive tweets differently based on unmeasured individual factors. Compared to a national representative sample used for Health Information National Trends Survey (HINTS, Kontos, Blake, Chou, & Prestin, 2014), the MTurk sample in this study was younger (about 10 years younger on average), less racially diverse (about 8% more Caucasians, 6% fewer Hispanics and Blacks), and more educated (about 9% more with a Bachelor's Degree). The characteristics of MTurk sample may present certain patterns of perceptions of health-related tweets.

Second, the study has “the missing denominator” problem shared by many published works using social media big data (Tufekci, 2014, p. 5); that is, the lack of data about the number of exposures to the CDC's original tweets and retweets. Without an adequate control for the viewing count data, our ability remains limited to make inferences about the effects of message- and network-features on information diffusion. Future research may use CDCs' number of followers and a typical exposure rate reported by Twitter as an estimate. Moreover, health messages studied were short given the word limit on Twitter. Twitter users possibly clicked and read the URLs included in the tweet messages before they made the decision about retweet. Therefore, the actual content led by URLs may influence the likelihood of retransmitting information, which was not captured by the current study.

Lastly, this study is also limited in its single-item measurement of some of the control variables including novelty and entertainment, which may explain the insignificant findings of the control variables in predicting diffusion outcomes. When defining novelty as being surprising and unusual, or offering a new argument deviating from existing schemata (Kim, 2015), studies have found novelty a significant message feature in increasing news retransmission. The single-item measure of perceived novelty in the survey may not have captured the rich meaning of it. Similarly, literature has referred entertainment as being interesting, funny, and extreme (Berger, 2014), which represents a broader spectrum of meanings.

6. Conclusion

This study contributed to the extant literature by examining the effects of both message and network features on health information diffusion on Twitter. This study also expanded scholarly focus on the size of diffusion (i.e., aggregated number of retweets) to structural pattern of diffusion (i.e., structural virality). The findings revealed that among message features examined, perceived negative emotion was a significant predictor of size and structural virality of diffusion; while perceived efficacy predicted diffusion size and perceived susceptibility predicted structural virality. The level of involvement of brokers was the only network feature that positively predicted structural virality. This study provided empirical evidence of distinct effects of message and network features on the two aspects of information diffusion. It also provided practical implications for public health organizations in terms of leveraging social media platforms for more effective information dissemination.

6.1. Note

1. The data collection method allowed us to collect 100% of original tweets posted by CDC in the period of data collection, but not 100% of retweets. When the number of retweets was smaller (= < 100), Twitter Search API was able to return all the retweets. When the number of retweets was greater (> 100), the search results had a few missing retweets when comparing with the number of retweets on CDC's timeline. However, the number of missing retweets was very small even among the most retweeted tweets (< 10). Therefore, we were confident that we had collected almost all the tweets that meet the search query. The rate limit was 180 requests per 15 min window for per-user authentication, and per request, we could ask for maximum 100 tweets. The grand total limit was 18,000 tweets/15 min. Moreover, we did not use Twitter's API feature such as RT to detect retweets. In our pilot data collection, we found that searching by exact tweet messages returned more accurate number of retweets than using RT, when comparing with the number of retweets shown on CDC's timeline.
2. Perceived response-efficacy and self-efficacy were correlated at 0.86, and therefore were not averaged into one index efficacy in the analysis.

Acknowledgement

This project is funded by Trifecta pilot grant award, Michigan State University.

Appendix A. Diffusion tree construction details

Here we describe the process of constructing a diffusion tree for a unique tweet. Trees are composed of one node for each retweeter and CDC, and each edge links a retweeter back to an inferred “parent”. To construct a diffusion tree for one unique tweet, that is, one diffusion event, we first identify potential “parents” of each user who retweeted the tweet. “Potential parents” are defined as individuals whose adoption of a tweet message appears in the focal retweeter's timeline prior to the focal retweeter's adoption (Goel et al., 2016). In other words, “potential parents” are the set of individuals who are likely to have exposed the user to the tweet. The second step is to infer a single parent from the “potential parents” of a given adoption. Based on the retweet time (i.e., the time at which each potential parent adopts the tweet content), we considered the one who tweeted most recently before the focal retweeter as the single “parent”.

Following this method, we could successfully trace back to the original author of each unique tweet – that is CDC – for all 425 unique tweets. In other words, CDC is the root at the very beginning of the diffusion cascade. However, in the diffusion trees of 27 unique tweets, there are some dyads and triads isolated from the major component. As noted by Goel et al. (2016), following the tree construction method, the parent need not be the original author of the tweet – that is CDC in this case. The reasons are that (1) users occasionally retweet content that did not appear in their timelines because they discovered it by browsing or searching on Twitter or a third party website, and (2) a followee who retweeted content by browsing or searching on Twitter or a third party website. By excluding these isolated dyads and triads, we still included 98.56% of all the retweeters in our dataset.

Appendix B. Details about trained coders

In addition to online survey evaluation to assess *perceived* message features, we had two trained coders content analyze the *expressed* message features of the 425 CDC Twitter messages. For each tweet, they coded *the presence* of the key message features studied in this paper: susceptibility, severity, response efficacy, self-efficacy, positive emotion (i.e., joy, love, hope, relief and compassion), negative emotion (i.e., fear, guilt, sadness, disgust and anger). Although we began with coding on a 5-point scale to be consistent with the scale used in assessing perceived message features, the two coders had a difficult time reaching an acceptable inter-coder reliability due to the limited words in each tweet and the latent nature of above-mentioned message features. Content analysis using human coders is a research technique for the objective description of the manifest content (Riffe, Lacy, & Fico, 2014, 3rd edition). This technique focuses on content's manifest meaning as opposed to latent “between-the-lines” meaning (Riffe et al., 2014, p. 30). Due to the limited words in each tweet, coding on a 5-point scale unavoidably invited the trained coders to speculate beyond manifested meaning.

When coding only the presence of key message features, in the third round of coding practice, intercoder reliability estimates (Cohen's kappa) ranged from 0.71 to 0.92 (M = 0.75) and the two coders reached an agreement of 91%. The final intercoder reliability estimates (Cohen's kappa) ranged from 0.73 to 0.95 (M = 0.83) and the coders reached an agreement of 93%.

Based on the good intercoder reliability, however, the two coders had some difficult time coding emotional features. Again, due to the short tweet messages, there were very limited words from which the two coders could use to infer emotion. For example, when a tweet contained words such as “risk” or “death”, it was coded into negative emotion “fear” because those words indicated “threat to one's physical or psychological self” (definition of fear, in Lazarus, 1991, p. 234). At the same time, “risk” and “death” are also key words for coding “susceptibility” and “severity” of a health issue. This resulted in high correlations between negative emotion and susceptibility and severity in human coding. However, with longer texts, the trained coders maybe more capable of coding emotional features. For example, Berger and Milkman (2012) had human coders successfully code emotions in New York Times articles.

The correlations between human coding (presence vs. absence) and survey rating (on 5-point scale) on efficacies and emotions ranged from 0.51 to 0.57. The correlations on susceptibility and severity were 0.34 and 0.35, relatively lower but statistically significant. It is worth noting that human coding and survey rating were used as methods to measure two distinct concepts: expressed and perceived message features (O'Keefe). Therefore, the positive but not high correlations are within our expectation.

Future study is encouraged to systematically examine and compare the advantages and disadvantages of several text analysis methods, including traditional human coders, computerized content analysis (e.g., LIWC), and online evaluation survey to crowdsource individual perceptions. A set of guidance could be made in terms of the appropriate message concepts to be measured (e.g., perceived vs. expressed, manifested vs. latent), types of texts to be applied (e.g., short tweets, long news articles), and other relevant issues.

References

Amabile, T. M. (1983). Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology, 19*, 146–156. https://doi.org/10.1016/0022-1031(83)90034-3.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. *Proceedings of the 21st international conference on world wide web* (pp. 519–528). ACM. https://doi.org/10.1145/2187836.2187907.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology, 24*, 586–607. https://

doi.org/10.1016/j.jcps.2014.05.002.
Berger, J., & Iyengar, R. (2013). Communication channels and word of mouth: How the medium shapes the message. *Journal of Consumer Research, 40*, 567–579. https://doi.org/10.1086/671345.
Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*, 192–205. https://doi.org/10.1509/jmr.10.0353.
Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2011). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*, 351–368. https://doi.org/10.1093/pan/mpr057.
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*, P1008. https://doi.org/10.1088/1742-5468/2008/10/P10008.
Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-mediated Communication, 13*, 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x.
Burt, R. S. (1992). *Structural holes: The social structure of competition.* Cambridge, MA: Harvard University Press.
Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data.* Cambridge, UK: Cambridge University Press.
Cappella, J. N., Kim, H. S., & Albarracín, D. (2015). Selection and transmission processes for information in the emerging media environment: Psychological motives and message characteristics. *Media Psychology, 18*, 396–424. https://doi.org/10.1080/15213269.2014.941112.
Centola, D. (2010). The spread of behavior in an online social network experiment. *Science, 329*, 1194–1197. https://doi.org/10.1126/science.1185231.
Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring user influence in Twitter: The million follower fallacy. *ICWSM, 10*(10–17), 30. https://doi.org/10.1016/j.ipm.2016.04.003.
Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world.* New York, NY: Cambridge University Press.
Ellison, N. B., Steinfield, C., & Lampe, C. (2011). Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society, 13*, 873–892. https://doi.org/10.1177/1461444810385389.
Figueiredo, C., Chen, W., & Azevedo, J. (2015). Central nodes and surprise in content selection in social networks. *Computers in Human Behavior, 51*, 382–392. https://doi.org/10.1016/j.chb.2015.04.070.
Fleming, L., & Waguespack, D. M. (2007). Brokerage, boundary spanning, and leadership in open innovation communities. *Organization Science, 18*, 165–180. https://doi.org/10.1287/orsc.1060.0242.
Galtung, J., & Ruge, M. H. (1965). The structure of foreign news the presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research, 2*, 64–90.
Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science, 62*, 180–196.
González-Bailón, S., Borge-Holthoefer, J., & Moreno, Y. (2013). Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist, 57*, 943–965.
Grandey, A. A. (2000). Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology, 5*, 95–110.
Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*, 1360–1380.
Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348–362.
Hansen, L. K., Arvidsson, A., Nielsen, F. A., Colleoni, E., & Etter, M. (2011). Good friends, bad news - affect and virality in Twitter. In J. J. Park, L. T. Yang, & C. Lee (Eds.). *Future information technology* (pp. 34–43). Berlin Heidelberg: Springer.
Harris, J. K., Mueller, N. L., & Snider, D. (2013). Social media adoption in local health departments nationwide. *American Journal of Public Health, 103*, 1700–1707.
Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing, 18*, 38–52.
John, N. A. (2012). Sharing and Web 2.0: The emergence of a keyword. *New Media & Society, 15*, 167–182.
Katz, E., & Lazarsfeld, P. F. (1955). *Personal Influence. The part played by people in the flow of mass communications.* New Brunswick, NJ: Transaction Publishers.
Kee, K. F., Sparks, L., Struppa, D. C., Mannucci, M. A., & Damiano, A. (2016). Information diffusion, Facebook clusters, and the simplicial model of social aggregation: A computational simulation of simplicial diffusers for community health interventions. *Health Communication, 31*, 385–399.
Kim, H. S. (2015). Attracting views and going viral: How message features and news-sharing channels affect health news diffusion: Attracting views and going viral. *Journal of Communication, 65*, 512–534.
Kissler, J., Herbert, C., Peyk, P., et al. (2007). Buzzwords: Early cortical responses to emotional words during reading. *Psychological Science, 18*(6), 475–480.
Knobloch-Westerwick, S., & Sarge, M. A. (2015). Impacts of exemplification and efficacy as characteristics of an online weight-loss message on selective exposure and subsequent weight-loss behavior. *Communication Research, 42*, 547–568.
Knobloch, S., Carpentier, F. D., & Zillmann, D. (2003). Effects of salience dimensions of informational utility on selective exposure to online news. *Journalism & Mass Communication Quarterly, 80*, 91–108.
Kontos, E., Blake, K. D., Chou, W. Y. S., & Prestin, A. (2014). Predictors of eHealth usage: Insights on the digital divide from the health information national Trends survey 2012. *Journal of Medical Internet Research, 16*.
Lazarus, R. S. (1991). *Emotion and adaptation.* NY: Oxford University Press.
Lou, T., & Tang, J. (2013). Mining structural hole spanners through information diffusion in social networks. *Proceedings of the 22nd international conference on world wide web* (pp. 825–836). .
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30–46.
McLaughlin, M. L., Hou, J., Meng, J., et al. (2016). Propagation of information about preexposure prophylaxis (PrEP) for HIV prevention through Twitter. *Health Communication, 31*, 998–1007.
Nahon, K., & Hemsley, J. (2013). *Going viral.* Cambridge, UK: Polity Press.
Newman, M. E. J. (2010). *Networks: An introduction.* Oxford, UK: Oxford University Press.
Onnela, J. P., Saramäki, J., Hyvönen, J., et al. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 7332–7336.
O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory, 13*, 251–274. https://doi.org/10.1111/j.1468-2885.2003.tb00292.x.
Riff, D., Lacy, S., & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research.* New York: Routledge.
Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review, 1*, 60–85.
Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). NY: Free Press.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.
Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media – sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*, 217–248.
Suh, B., Hong, L., Pirolli, P., et al. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. *Proceedings of the IEEE second international conference on social computing* (pp. 177–184). .
Sysomos (2010). *An in-depth report at Twitter's growth 2010, compared with 2009.* Retrieved March 29, 2016, from http://sysomos.com/inside-twitter/twitter-statistics.
Thelwall, M., Buckley, K., et al. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology, 62*, 406–418.
Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Proceedings of the 8th international AAAI conference on weblogs and social media* (pp. 505–514). .
Valente, T. W. (1995). *Network models of the diffusion of innovations.* Cresskill, NJ: Hampton Press.
Valente, T. W., & Fujimoto, K. (2010). Bridging: Locating critical connectors in a network. *Social Networks, 32*, 212–220.
Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*, 1146–1151.
Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research, 34*, 441–458.
Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports, 3*, 2522.
Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs, 59*, 329–349.
Witte, K., Cameron, K. A., Mckeon, J. K., et al. (1996). Predicting risk behaviors: Development and validation of a diagnostic scale. *Journal of Health Communication, 1*, 317–342.
Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. *Proceedings of the 4th international AAAI conference on weblogs and social media* (pp. 355–358). .
Zhao, J., Wu, J., Feng, X., et al. (2012). Information propagation in online social networks: A tie-strength perspective. *Knowledge and Information Systems, 32*, 589–608.

Jingbo Meng (PhD, University of Southern California) is an Assistant Professor in Department of Communication at Michigan State University. She is generally interested in integrating network theories into classic behavioral change models and information diffusion theories. Specifically, her research focuses on the formation and influence of social networks on health promotion and information diffusion on social media.

Wei Peng (PhD, University of Southern California) is an Associate Professor at the Department of Media and Information, Michigan State University. Her research focuses on the psychological and social mechanisms of behavior change and their application in the design of interactive media for health and wellness promotion.

Pang-Ning Tan (PhD, University of Minnesota) is a Professor in the department of Computer Science and Engineering at Michigan State University. He has published over 100 peer-reviewed articles in conferences, journals, and book chapters in the areas of data mining and machine learning.

Wuyu Liu (PhD, Michigan State University) is an Assistant Professor in Department of Communication at Bellarmine University.

Ying Cheng is a doctoral candidate in the Department of Communication at Michigan State University.

Arram Bae is a doctoral student in the Department of Communication at Michigan State University.