What makes cancer information viral on social media?

Wang, Xiaohui; Chen, Liang; Shi, Jingyuan; Peng, Tai Quan

Published in:
Computers in Human Behavior

DOI:
10.1016/j.chb.2018.12.024

Published: 01/04/2019

Document Version:
Peer reviewed version

What Makes Cancer Information Viral on Social Media?

Xiaohui Wang[a], Liang Chen[b], Jingyuan Shi[c], Tai-Quan Peng[d]

a. Department of Journalism, Hong Kong Baptist University, Hong Kong
b. Lab for Big Data and Public Communication, School of Communication and Design, Sun Yat-sen University, Guangzhou, China
c. Department of Communication, Hong Kong Baptist University, Hong Kong
d. Department of Communication, Michigan State University, USA

Wang, X., Chen, L., Shi, J., & Peng, T-Q (2019). What makes cancer information viral on social media? *Computers in Human Behavior*, *93*, 149–156.
https://doi.org/10.1016/j.chb.2018.12.024

Abstract

Although critical in practice, as well as prevalent, limited effort has gone into understanding the driving factors behind the diffusion of cancer information on social media. This study seeks to comprehensively examine both the content and sender factors that determine cancer information diffusion. Specifically, a multidimensional measure is proposed to capture diffusion characteristics of cancer information, including the scale of information diffusion, structural virality of information diffusion, and the engagement in cancer discussion. A conceptual framework consisting of content and sender factors is proposed based on the heuristic-systematic model to account for cancer information diffusion. Through analyzing information diffusion networks of cancer-related information on Weibo, the results revealed that sender factors played determining roles in affecting information diffusion. The content factors also significantly predicted information diffusion. The factors driving engagement are substantially different from those of scale and virality of information diffusion. Implications for both information diffusion theories and cancer education on social media are discussed.

Keywords: Information diffusion, Cancer, Sender factors, Content factors, Social media

What Makes Cancer Information Viral on Social Media?

Fostered by the ubiquitous information technologies, social media platforms, such as Twitter, Facebook, YouTube, and Weibo (a Twitter-like social medium in China), have rapidly developed as an important source for information publishing and sharing. Social media has not only expanded the availability, timeliness, and accessibility to online information but also changed the methods of information dissemination and acceptance and, ultimately, the end-user behaviors (Kim, Hou, Han, & Himelboim, 2016; Villagran, 2011).

Increasing attention from both academia and industry has explored how to leverage such platforms for greater societal benefits, especially in health promotion and cancer education (e.g., Heo, Chun, Lee, & Woo, 2018; Loeb, Katz, Langford, Byrne, & Ciprut, 2018; Watson, 2018). A series of social media health campaigns has been run to effectively improve the public's cancer prevention behaviors, such as breast self-examination (Chen & Yang, 2018). Cancer is a major global public health issue. In 2016, approximately nine million people were estimated to have died from the various forms of cancer all over the world. As the most populous country, China is facing a huge burden of cancer. In China, cancer is the leading cause of death, with 4.3 million new cancer cases and 2.8 million cancer deaths estimated to occur each year (American Cancer Society, 2017).

Social media are a critical source for cancer prevention. For example, microblogs could serve as a source for up-to-date cancer information about prevention, diagnosis, and treatment (O'Neill, 2017; Loeb et al., 2018), and provide information with emotional and informational support for cancer patients and their caregivers (Shi, Chen, Su, & Chen, 2018). Although cancer information is prevalent on social media, research has shown that not all information goes viral and spreads throughout social media; instead, most information languishes in obscurity (Susarla, Oh, & Tan, 2012). Indeed, modeling online information diffusion has proven to be a challenging task (Yang & Leskovec, 2010). Previous research efforts have explored the factors related to the popularity of certain information on social media and made significant progress in improving the model of driving factors (Kim et al., 2016). However, a systematic model of the driving factors behind information diffusion remains underexplored. Therefore, in the current study, we take cancer information as an example and investigate why certain pieces of cancer content go viral on a Chinese social media platform, Weibo. Similar to Twitter, Weibo users can response to others' posts by liking, commenting upon, and sharing the content. In the current study, we aim to investigate what are the driving factors behind the dissemination of cancer-relation information on Weibo?

The current research contributes to the literature in two aspects: (1) developing a multidimensional measure of information diffusion on social media by including scale, structural virality, and engagement; and (2) investigating the underlying mechanism of online information diffusion using the content, as well as sender factors, in social media posts. The findings of this study provide implications for both information diffusion theories and cancer education on social media.

The Multi-Dimensional Nature of Information Diffusion

Information diffusion (also known as information spread, transmission, or propagation) is a long-standing research domain and has attracted research interest from many fields (Guille, Hacid, Favre, & Zighed, 2013). Modeling the information diffusion process is of outstanding interest since it provides the fundamentals to understand and manage the dynamics of content on social media, such as minimizing rumor propagation or optimizing the performance of health campaigns (Guille et al., 2013).

Scholars have studied online information diffusion from various perspectives. Some have mapped the information diffusion network (Cheng, Adamic, Kleinberg, & Leskovec, 2016; Goel, Watts, & Goldstein, 2012; Himelboim & Han, 2014) and examined the driving factors of information diffusion on Twitter (e.g., Bakshy, Hofman, Mason, & Watts, 2011; Suh, Hong, Pirolli, & Chi, 2010) and Weibo (e.g., Liu, Liu, & Li, 2012). Still others have explored the driving factors of the popularity of health information on social media (Briones, Nan, Madden, & Waks, 2012; Kim et al., 2016; Ma, Sun, & Cong, 2013; Petrovic, Osborne, & Lavrenko, 2011). However, a major limitation in previous research is the operationalization of information diffusion. The information diffusion in previous research was defined vaguely and measured generally, where its multiple dimensions have rarely been explored.

Table 1 presents an incomplete summary of the measurements proposed in previous empirical studies of information diffusion. The conceptualizations and operationalizations of the characteristics of online diffusion can be mainly categorized into three dimensions: diffusion size, diffusion speed, and structural characteristics of diffusion networks. Diffusion size represents the total number of audiences involved in the spreading of the information, which is also known as popularity, and breadth [measured as the total number of retweets or likes (Briones et al., 2012; Kim et al., 2016; Ma et al., 2013)] and scale [number of first-degree retweets (Yang & Counts, 2010)]. Diffusion speed is the time interval between an information diffusion behavior; that is, how fast a piece of content is spread (Yang & Counts, 2010). Finally, network structure identifies the structural features of the information diffusion network. The structure of the information diffusion network is measured with network indices such as range and depth [equal to the longest geodesic distance between the original tweet and its retweets (Wasserman & Faust, 1994)] and structural virality [measured by the mean of the distance from the focal nodes to all the other nodes in the network (Goel, Anderson, Hofman, & Watts, 2015)].

Collectively, most previous studies have viewed only retweeting as the behavioral outcome of information diffusion on social media. However, commenting should also be considered as a behavioral outcome. Retweeting and commenting both involve users in the dissemination of the information on social media (Boyd, Golder, & Lotan, 2010). Thus, in this study, we integrated retweeting, as well as commenting processes on social media, to develop a multidimensional construct of information diffusion. Specifically, we use scale and structural virality measuring the structure of the diffusion process and number of comments measuring the user's engagement in the diffusion process. Each of the subdimensions will be reviewed in detail in the following sections.

Table 1

List of Measurements and Subconstructs of Online Information Diffusion in the Literature.

		Measurements of Information Diffusion			
Publication	**Platform**	**Diffusion Size**	**Diffusion Speed**	**Network Structure**	**Others**
Cha, Mislove, and Gummadi (2009)	Flickr	√ (likes)	√		
Hong and Davison (2010)	Twitter	√ (retweets)			
Suh et al. (2010)	Twitter				Retweet Rate
Yang and Counts (2010)	Twitter	√ (1-degree neighbors)	√	√ (range)	
Petrovic et al. (2011)	Twitter				Retweet or not
Bakshy, Rosenn, Marlow, and Adamic (2012)	Facebook				Repost or not
Liu et al., (2012)	Weibo	√ (retweets)			
Stieglitz and Dang-Xuan (2013)	Twitter	√ (retweets)	√		

Starbird and Palen (2012)	Twitter	√ (retweets)	√		
Weng, Menczer, and Ahn (2013)	Twitter	√ (adoptions)			
Goel et al. (2015)	Twitter	√ (retweets),		√ (Structural Virality)	
Zhang and Peng (2015)	Weibo	√ (retweets)	√	√ (range)	
Kim et al. (2016)	Twitter				Retweet or not
Meng et al. (2018)	Twitter	√ (retweets)		√ (Structural Virality)	

Scale and Structural Virality in Retweet Network

When a social media user retweets a post, it means that the user has paid attention to the tweet, wanted to broadcast this message to his/her own followers, or intended to publicly show agreement or disagreement with someone (Boyd et al., 2010). The concepts of scale and structural virality stem from two major, but distinctive, information diffusion modes in the retweet network.

First, broadcasting is one diffusion mode, which refers to a large number of individuals receiving the information directly from the same social media user. The information spread quickly because of the large number of subscribers of the original user, which could be an account belonging to traditional media, governments, or celebrities (Goel et al., 2015). This process is a “one-to-many” process where a social media user influences the mass audience around a topic (Morris & Ogan, 1996).

The second diffusion mode on social media is contagion. During this process, the content attained its popularity through a process of person-to-person influence, analogous to the spread of a virus (Anderson & May, 1992; Dodds & Watts, 2004). A piece of information in social media may originate from a grassroots user with a small number of followers. Although his or her information may not reach a massive audience initially, it could be accepted and disseminated by his or her close neighbors, and finally reach many individuals through multiple generations of peer-to-peer disseminations; i.e., the process of "one-to-few" (Morris & Ogan, 1996).

Figure 1 illustrates these two modes of information diffusion. The former is broadcast, referring to a star-like network structure where the burst of diffusion is mainly due to a single influential node. The latter is contagion, which comprises a multigenerational branching structure in which nodes directly influence only a few others (Dodds & Watts, 2004). Corresponding to the two modes of information diffusion, the measurement of information diffusion can be further decomposed into scale and structural virality. Scale refers to the number of audiences involved in directly spreading the information seed (Yang & Counts, 2010), and structural virality refers to the divergent branches of the information diffusion structure (Goel et al., 2015). A dispersed information structure means that the information is more likely to be shared with more heterogeneous communities and with less information overload, and thus can be seen as a more positive outcome of information diffusion.

A major limitation in previous research is that classical diffusion studies typically had access only to aggregate diffusion data, such as popularity. Part of the reason could be that the data for the structural properties are difficult to obtain (Goel et al., 2015). In spite of a large number of theoretical and empirical studies on the diffusion of information, the structure of information diffusion on social media and its driving force remains underexplored. Understanding the structure of the diffusion requires reconstructing the full topology of the diffusion paths. Such action is lacking in online cancer information diffusion studies.

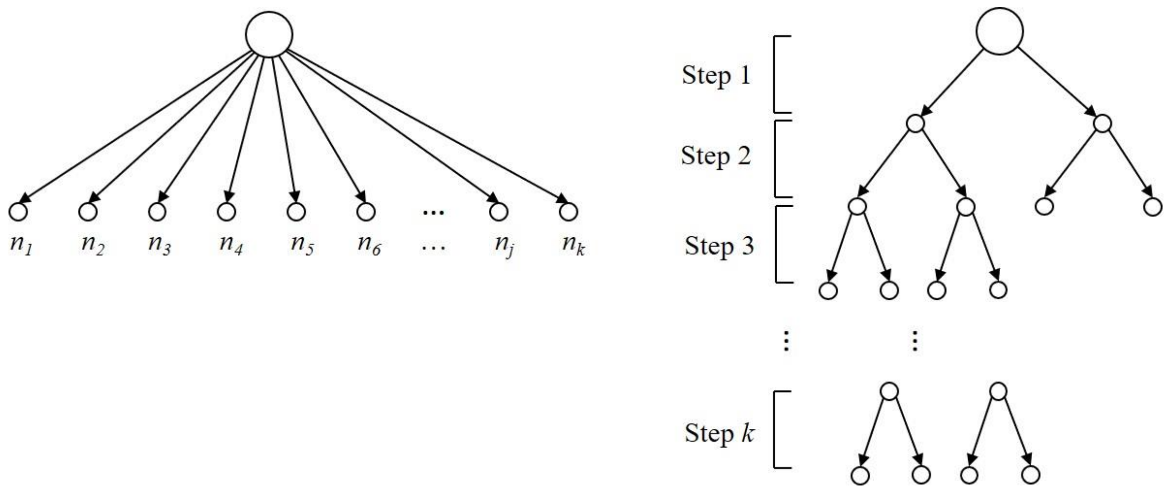

Figure 1. A schematic illustration of the broadcast and contagion effects.

Commenting as Engagement in Information Diffusion

Although most of the previous diffusion research has investigated the retweeting outcome of information diffusion (Table 1), few have considered commenting as a behavioral outcome of a diffusion process. Retweeting means the audiences have listened to the original post, whereas commenting involves more audiences' cognitive effort, given that commenting generally requires editing new content that adds value to the original post (Boyd et al., 2010). Comments from the audience may alter the interpretation of the original content. In the current study, we argue that commenting is a behavior where the audience is engaging in information diffusion and should be treated separately and differently from retweeting.

In addition, information diffusion on social media is not simply to share information to others but also to bring new people into a particular thread anticipating that they will engage in the topic. Commenting is the means through which conversations develop on social media. When commenting on a post, users may validate the original content, provide additional arguments, or argue against the original content; thus, comments could shape how the original content is processed and may add new value to the interpretation of the original post (Lee & Tandoc, 2017). Moreover, by engaging in the discussion, the commenting function on social media also initiates and promotes conversations between various users (Boyd et al., 2010). Ultimately, this commenting behavior by numerous users creates “buzz,” in which the interactions among audiences amplifies the dissemination of the original message (Thomas, 2006). Thus, commenting should also be considered as a behavioral outcome of the engagement in the process of information diffusion.

Collectively, we propose a multidimensional construct of information diffusion that integrates retweeting, as well as commenting processes, on social media. First, scale and structural virality of a retweet network allow us to quantify the diffusion outcomes induced by broadcast and contagion diffusion mechanisms, providing more fine-grained distinctions of the measurements of the retweet network. In addition, understanding commenting behavior provides detailed knowledge about how a piece of content goes viral in social media. Thus, we developed a three-dimensional measurement of information diffusion as follows: the scale and structural virality measuring the size and structure of the retweet network, and the number of comments measuring users’ engagement in information diffusion.

Factors Driving Cancer Information Diffusion

Given the important roles of retweeting and commenting in spreading information and promoting interpersonal discussion, it is important to examine the factors driving such behaviors. From this, one could develop a theoretical framework for online information diffusion and enable practitioners in strategic campaign design. Although many efforts have been investigated to model the information diffusion process, the underlying mechanism of information spread remains underexplored.

Information processing theories have suggested that message content, as well as source characteristics, influence how individuals process the information (Chaiken, 1980; Petty & Cacioppo, 1986). Information diffusion is a complex sociocognitive process (Casterline, 2001), which can be viewed as a behavioral decision that is made after assessing the particular information (Liu et al., 2012). As such, the heuristic-systematic model (HSM) of information processing (Chaiken, 1980), one of the prevalent information processing models, can be used to explain information diffusion. The HSM posits that a message can influence people's attitudes and behaviors in two different ways: systematic and heuristic processing. The former focuses on the message itself, such as argument quality and content elements, which requires recipients to use high-cognitive efforts to elaborate information. The latter focuses on certain cues, such as its source and peer recognition, which enables individuals to adopt heuristic and simple decision rules to quickly form judgments (Chaiken, 1980; Chaiken & Eagly, 1989).

Systematic Factors: Content Factors

According to the HSM (Chaiken, 1980; Chaiken & Eagly, 1989), the message itself, such as argument quality and content elements, could determine how individuals process information. Previous efforts investigating the content factors of the diffusion of information in social media focused on signal characteristics such as the sentiment of the tweets, number of URLs, mentions, multimedia, and quality and length of the messages (e.g., Liu et al., 2012; Stieglitz & Dang-Xuan, 2013; Srivastava, Saks, Weed, & Atkins, 2018; Suh et al., 2010). The results from these studies suggest a framework of the driving factors of information diffusion in general settings. However, a context-specific model with systematic factors is lacking, which could offer definite guidelines for the design of campaigns on social media. Especially in the field of cancer prevention, cancer-related content factors, such as topics in the messages and discrete emotions, are more important but are ignored in the current literature.

Examining the association between topics and diffusion outcomes could offer insight into the foci of public attention and which topic is more prevalent and more widely spread among users. Hong, Dan, and Davison (2011) is one of the few studies investigating the relationship between the topic and the popularity of messages. They adopted an automatic text mining technique to detect popular topics in tweets and found it contributes to the information diffusion. However, topics generated by this bottom-up approach were general categories with vaguely defined boundaries. Instead, in this study, we rely on a top-down approach, manually coding cancer-related topics from the tweets. We would like to examine what cancer topics are the focus of public attention. Understanding this could provide meaningful insight for policy making and health campaigns.

Emotion in the messages is also suggested as an important factor that may affect the diffusion process (Briones et al., 2012; Stieglitz & Dang-Xuan, 2013). As part of communication, online content often conveys emotional information or sentiment. The results from social media studies indicate that the affective polarity of the online messages (positive, negative, or both) could trigger a higher level of cognitive engagement and arousal, which may in turn affect information sharing (Berger & Milkman, 2012). Importantly, however, the spreading of emotional content may be driven by more than just polarity. In addition to being positive or negative, emotions also differ on the level of physiological arousal they evoke (Smith & Ellsworth, 1985), and these differences may shape information diffusion (Berger & Milkman, 2012). Thus, this research goes beyond psychological valence, to examine how discrete emotions drive information diffusion.

In the context of cancer, fear and hope as discrete emotions are usually discussed. According to the terror management theory, humans are instinctively driven toward survival and continued existence, while at the same time having knowledge of their own mortality (Solomon, Greenberg, & Pyszczynski, 1991). As a result, when talking about cancer that a topic highly associated with mortality, individuals usually experience a death fear. However, hope has been considered as a particularly effective buffer against death fear (Wink & Scott, 2005). As such, online cancer information may deliver fear and hope, which could be crucial elements in information diffusion.

Heuristic Cues: Sender Factors

In addition to the message content, the sender is the other important factor affecting whether and how a tweet goes viral on social media. According to the HSM (Chaiken & Eagly, 1989), individuals will rely on heuristic cues for information processing if they lack motivation or the ability to scrutinize the content. In the context of cancer information seeking online, social media users who may have little professional and medical knowledge are highly likely to rely on heuristic cues for information processing. One of the most important source characteristics is credibility (Bakshy et al., 2011; Liu et al., 2012; Watts & Zhang, 2008), defined as the extent to which an information source is perceived to be believable, competent, and trustworthy (Petty & Cacioppo, 1986).

First, the numbers of followers (i.e., audience) and followees (i.e., information sources) have been regarded as the indicators of social influential power and credibility of a social media user (Westerman, Spence, & Van Der Heide, 2012). They directly affect the dissemination of information on social media (Goldenberg, Han, Lehmann, & Hong, 2009; Suh et al., 2010). From a network perspective, social media users with large numbers of followers and followees are centrally and well connected. As such, they have a great ability to transmit information to others (Lahuerta-Otero & Cordero-Gutiérrez, 2016). Thus, we expect that numbers of followers and followees are vital determinants of online information diffusion.

The second indicator of a sender's credibility is its account characteristics. User profiles such as a verified account and whether the account is identified as an organizational account or a professional account, is open to the public, and could indicate the credibility of the users (Zhang, Peng, Zhang, Wang, & Zhu, 2014). These characteristics might gain authentic value for the content and thus facilitate the spread of such content. The effect of verifying status and account type has been suggested in the literature (Himelboim & Han, 2014; Liu et al., 2012). In this study, we divide the account type into three categories tailored to the context of cancer information: medical-related professionals/organizations, nonmedical organizations, and grassroots. This context-specific categorization of users is expected to account for the variance in cancer information diffusion online.

In summary, according to the HSM, we aim to investigate the underlying factors of cancer information diffusion on social media. The proposed predictors include systematic factors (i.e., topics, sentiment, and discrete emotions) and heuristic factors (i.e., numbers of followers and followees, verifying status, and account type).

Method

To investigate cancer information diffusion in social media, we used social network analysis, content analysis, sentiment analysis, and multivariate analysis of variance (MANOVA) to address our research questions. Specifically, social network analysis presents an overview of the retweeting network structure; i.e., the magnitude and structure of the cancer information diffusion. Quantitative content analysis unobtrusively identifies the pattern of message content and characteristics of the senders. This information is then used in the MANOVA analysis.

Data Collection

We used the keyword "cancer" to locate cancer-related posts on Weibo, a leading and representative microblog platform in China, for the year 2015-2016. We randomly selected seven weeks and captured the cancer-related posts published in each week. All the posts were extracted using the Python Web Crawler. The dataset contained the following meta-data for the posts: posting time, number of retweets, number of comments, number of likes, and the account name of the sender. The profile information for the sender was also recorded, which included: number of followees, number of followers, number of history posts published on Weibo, and whether it is a verified account. The dataset was manually cleaned before the analysis. Specifically, commercial advertisements and messages not related to cancer were eliminated from the sample (2,038 posts were eliminated). All posts in this dataset were written in Chinese. The final dataset consisted of 14,616 posts. The second stage of data collection was mapping the retweet networks. We used all the posts in the dataset as seeds to map the topology of the retweet networks. To minimize the potential harm to participants, confidentiality was ensured through replacing the usernames with pseudonyms and paraphrasing the messages published in this manuscript to avoid identification of the users.

Manual Content Analysis

We relied on human coders to classify the extent to which content and sender exhibited specific characteristics because automated coding systems were not available for these characteristics. Such characteristics include the following: (1) whether the tweet is talking about cancer prevention (i.e., a prevention message) or talking about personal experience of cancer (i.e., a personal experience message), (2) whether the tweet contains fear or hope emotion, and (3) whether the sender of the tweet is an account that belongs to medical professional/organizations, nonmedical organizations, or grassroots/ordinary users. The coders were blind to the study purposes. Specifically, two trained coders who are native Chinese speakers independently coded 10.36% (n = 1,726) of the tweets to establish the intercoder reliability. Krippendorff's alpha was reported as .90 (individual experience), .89 (prevention), .85 (fear), .82 (hope), and .95 (account type), which were acceptable. Discrepancies were resolved, and the coding rules were established via discussion to avoid ambiguities in word meanings, category definitions, and coding instructions. The remaining messages were then split in half and separately coded by the two coders.

Automatic Sentiment Analysis

We conducted sentiment analysis to measure whether the tweet conveys a generally positive or negative sentiment. Sentiment analysis mines individuals' opinions, attitudes, and emotions from written language. This form of analysis is one of the most active research areas in text mining because opinions, or sentiments, are central to human activities and are key influencers of an individual's future behaviors and decisions. In this study, the sentiment score for each tweet was calculated using dictionary-based sentiment analysis and validated by manual coding[1].

Measurements of Information Diffusion and Driving Factors

For each post, a retweet network of it was mapped. We calculated each retweet network metrics, including the number of 1-degree neighbor of the sender and the structural virality for the tweet. The number of 1-degree users involved in the retweet network measured the scale of the information diffusion. Structural virality is proposed as an additional index of the information diffusion, measured by the mean of the distance from the focal nodes to all the other nodes in the diffusion network. Finally, engagement is measured by the number of comments received for each post. Moreover, some diffusion indices were also calculated in this study for descriptive purposes, such as total number of retweets and range of the diffusion network, which is equal to the longest geodesic distance between the original tweet and its retweets.

For the measurement of driving factors of information diffusion, number of followers, number of followees, and verifying status were directly crawled from the user profile; account type (medical, organizational, and grassroots), topics (prevention-related and personal experience-related), and emotions (hope and fear) were from the manual coding results; and sentiment scores were generated by automatic sentiment analysis. We also included control variables suggested by the previous literature, including whether a message contains a URL, hashtag, or mention, and the number of historical tweets. All the constructs and measurements are listed in Table 2.

Table 2

The Measurement of Constructs.

Constructs	Measurement
Dependent Variables	
Scale	The number of 1-degree neighbors in the retweet network
Structural Virality	The mean of the distance from the focal nodes to all the other nodes in the retweet network
Engagement	Number of comments received
Content Factors	
Emotion: Fear	Whether the tweet contains fear emotion
Emotion: Hope	Whether the tweet contains hope emotion
Sentiment	Whether the tweet conveys positive or negative emotion; larger absolute value indicates more affective words in the tweet
Topic: Experience	Whether the tweet is personal experience-related
Topic: Prevention	Whether the tweet is cancer prevention-related
Sender Factors	
Medical Account	Medical related professionals/organizations
Organizations	Nonmedical organizations, i.e., mainly governments and news media
Grassroots	Ordinary users
Verifying status	Whether the account is verified or not
Followers	Number of fans/audiences
Followees	Number of followings/information sources
Control Factors	
URL	Whether the tweet contains URL
Hashtag	Whether the tweet contains hashtag (#)
Mention	Whether the tweet contains mentions (@)
Activeness	Number of history tweets published on Weibo

Results

Descriptive Analysis

A summary of the retweet networks showed that most of the cancer messages are not retweeted (82.5%). Among those who received retweets, the majority of message cascades is tiny and terminates at the first generation (71.6%). However, some tweets can be very popular, and their cascades become very large: the most popular post received 8,033 retweets. It seems that although the vast majority of the cancer-related tweets vanish, large cascades do occur, at an extremely low rate. Similar results were found for the number of comments. Most of the cancer messages received no comments (73.7%), while some messages aroused intense discussion among audience (number of comments = 10,897). This result confirms the finding in previous studies that large cascades are very rare events (Bakshy et al., 2011; Goel et al., 2015; Goel et al., 2012).

The maximum range of the diffusion path is 15, which is smaller than reported in other studies [Zhang and Peng (2015) found a maximum of 28; Goel et al. (2015) found a maximum range of 34]. The small value for range suggested that the depth of the cancer information diffusion network is short compared with other messages, which may limit the diversity of the audiences for cancer information. This short range also suggests that cancer information promotion may face more difficulties than other topics.

A large amount of cancer information was personal experience and cancer prevention-related messages. In the 14,616 messages, 33.9% (n = 4,958) messages described personal experience and 33.1% (n = 4,547) messages contained cancer prevention information. Only a few cancer messages explicitly expressed fear or hope emotion (350 and 72 out of 14,616). The categorization of users according to their Weibo profile revealed that approximately 8.7% were medical-related professionals/organizations and 14.3% of the users were nonmedical organizations, such as government and news media. The rest were grassroots (77.0%). Among those who received retweets, 14.0% were medical accounts, 33.4% were organizational accounts, and 40.1% were ordinary users. Obviously, nonmedical organizational accounts received more retweets compared to other types of accounts.

Examining Broadcasting and Virality

This study was also conducted to investigate whether the broadcast and contagion effects work together in the context of cancer information diffusion. We tested the bivariate correlation between scale and structural virality among posts that received retweets (n = 2,923). Consistent with the findings of Goel et al. (2015), the correlation between scale and structural virality in our data is significant and weak ($r = .19$, $p < .001$), showing that the relationship between diffusion size and network structure is positive but noisy. This result suggests that structural virality is a concept that represents the diffusion topology in a different dimension from the scale measures. Structural virality effectively quantifies different diffusion features by measuring the structure of the diffusion process that cannot be captured by a cumulative index. On the other hand, the index of structural virality at each level of diffusion size is diverse. For instance, for a given diffusion size level (size of 95 to 105 retweets, for instance), the diffusion depth is between 2 to 13. As seen in Figure 3, information diffusion at the given level of diffusion size can be driven purely by broadcast, in the sense that all the users receive a message from a single source (top left), and high structural virality, in that the message diffuses through multiple generations and branches (bottom right). The combinations of the two modes indicated that information diffusion is driven by a mix of the broadcast and contagion mechanisms.

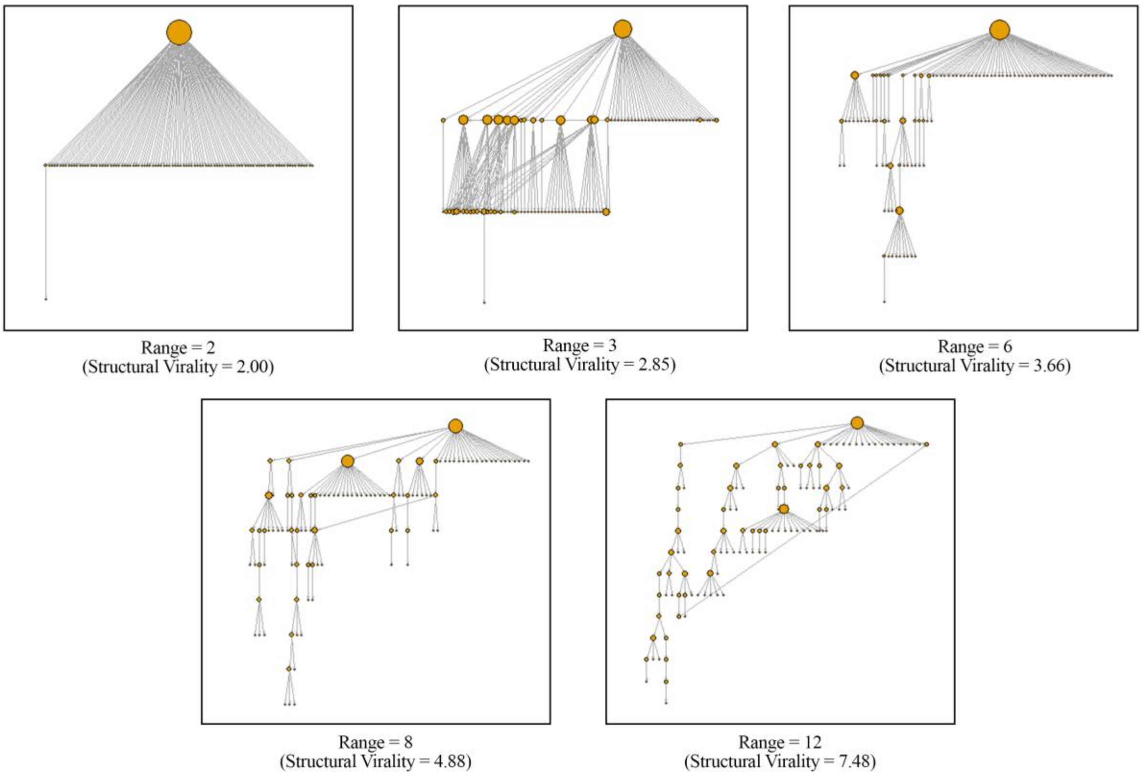

Figure 3. an example of the mix of broadcasting and contagion effects (diffusion size 95-105, diffusion depth ranging from 2 to 12, node size was weighted by indegree).

Examining Driving Factors of Information Diffusion

To determine whether both sender and content factors were positively associated with information diffusion and whether each accounted for unique variance in information diffusion, we used MANOVA to explain the three correlated information diffusion concepts, including scale, structural virality, and engagement. First, the control factors, including URL, hashtag, mention in the content, and activeness of the sender, were entered in the model. The second block were sender factors, including account type, number of followers, number of followees, and verifying status, while the third block were the content factors including discrete emotion (fear and hope), sentiment, and topic (prevention and experience).

By comparing the blocks and models, we found that the content factors only explained a small portion of the variance of scale (4.8%), structural virality (5.2%), and engagement (11.8%). The sender factors explained an additional 39.2% of the variance in scale, 33.6% of the structural virality, and 22.2% of the engagement.

As shown in Table 3, in terms of explaining the scale of the retweet network, content and sender factors explained a significant portion of the variance in scale. Tweets expressing hope emotion, without fear emotion, with higher sentiment, experience-relevant content, and prevention-relevant content were also positively and significantly associated with the scale of the retweet network. Tweets from senders that are medical accounts or nonmedical organizations, with a larger number of followers, a smaller number of follows, and unverified users were more likely to be popular on Weibo.

Similarly, in terms of explaining the structural virality of the retweet network, sentiment, experience-relevant content, prevention-relevant content, and mention were also positively and significantly associated with the virality of the retweet network. The effect from hope emotion on virality was significant, while fear emotion was significant but negative. In the block of sender factors, MANOVA analysis found that medical accounts, nonmedical organizations, number of followers, number of follows, and verifying status were significant predictors of structural virality.

Factors influencing engagement were different from predictors of retweet scale and structural virality. We found that fear emotion, hope emotion, personal experience-relevant content, and number of followers were positively and significantly related to the number of comments received. In contrast, the effect of sentiment, medical accounts, nonmedical organizations, verified account, and number of followees were found to be significant but negative. The effect of prevention-related content in tweets was nonsignificant.

Table 3

MANOVA Results of Diffusion Scale, Structural Virality, and Engagement (n = 14,616).

	Scale		Structural Virality		Engagement	
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	.551 ***	.128*	.754 ***	.269**	1.113 ***	.530 ***
Control Factors						
URLs	.109***	.003	.109***	.005	-.012	-.077 ***
Hashtags	.073***	-.011	.084***	.001	.026 **	-.024 **
Mentions	.035***	.040***	.054***	.058***	.033 **	.036 ***
Activeness	.105***	-.151***	.118***	-.113***	.067 ***	-.099 ***
Content Factors						
Fear [0,1]	-.021*	-.016*	-.019*	-.015*	.059 ***	.063 ***
Hope [0,1]	.077***	.041***	.061***	.028***	.116 ***	.084 ***
Sentiment	.072***	.014*	.075***	.018*	-.005	-.029 **
Experience [0,1]	.153***	.089***	.159***	.107***	.311 ***	.251 ***
Prevention [0,1]	.081***	.040***	.082***	.044***	.037 ***	.007
Sender Factors						
Followee		-.075***		-.069***		-.041 ***
Follower		.783***		.678***		.640 ***
Verifying Status [0,1]		-.095***		-.044***		-.106 ***
Medical Account [0 = Grassroots]		.021**		.042***		-.038 ***
Nonmedical Organization Account [0 = Grassroots]		.028**		.046***		-.049 ***
Adjusted R^2	.059 ***	.451***	.067 ***	.403***	.122 ***	.344 ***
ΔR^2		.392***		.336***		.222 ***

Note: **p* < .05; ***p* < .01; ****p* < .001

Discussion

This study was initiated to examine the driving factors spreading of cancer-related information on social media. The cancer-related tweets on Weibo were sampled over a span of one year from 2015 to 2016, and their retweet networks were mapped to address our research objectives. We argue that information diffusion on social media is not a simple and straightforward concept that can be fully captured using a single continuum scale, but rather a complex concept that involves three separate dimensions: scale, structural virality, and engagement. A model consisting of systematic and heuristic factors is proposed to explain the three dimensions of information diffusion. MANOVA results revealed that the sender factors played determining roles in affecting information diffusion. The content factors were also significantly associated with information diffusion. However, their effects were relatively marginal compared to the sender effects.

First, this study induces a multiple-dimensional measure of online information diffusion, by integrating two behavioral outcomes on social media: retweeting and commenting. The three subdimensions include scale, structural virality, and engagement. Although the current analysis is conducted based on the data from Weibo, the research approach and method can be applied to the general setting. Large cascades occur in cancer information diffusion, ranging from a fully broadcast structure to a highly viral structure. Our descriptive result demonstrates the mix effect of these two mechanisms. The relatively low correlation between scale and structural virality also highlights the limitations in the extant studies. Using only popularity of the retweets as the index for information diffusion is not enough for the understanding of how content goes viral on social media. A more comprehensive measurement of information diffusion should be multiple-dimensional in nature, as in this study.

The factors, such as numbers of followers and followees, verifying status, and topic, are the major determinants of online information diffusion, suggesting that these factors are essential for health campaign design. Interestingly, verifying status is negatively related to scale and virality of retweets. This result suggests that, in accounting for the effect of account type, number of followers, and number of followees, tweets from unverified accounts are more likely to be accepted by the audience. One possible explanation could be that most verified accounts are news media or agencies that usually play a strong propaganda role for the Chinese government, especially in the topic of health. Previous studies have indicated that compared with propaganda media, Chinese social media users are more likely to comment and retweet the content generated at the grassroots level (Chen, Wang, & Peng, 2018; Shi & Salmon, 2018).

Emotion in the content also affects the retweeting of cancer information. By zooming into specific discrete emotions, only hope is a significant predictor, while fear has no significant effect. This outcome may be because most Chinese people are well aware of the threat of cancer in general. Moreover, according to the terror management theory, hope is an effective buffer against death fear (Wink & Scott, 2005). Thus, most people tend to provide and diffuse positive or supportive emotionally messages on social media, which could be helpful for cancer patients and their caregivers (Shi et al., 2018). Although the effect of emotion in information participation and sharing has been emphasized in previous literature (Berger & Milkman, 2012; Stieglitz & Dang-Xuan, 2013), little is known about which emotion works in the context of cancer information. In fact, fear has been a common emotion in cancer communication messages designed by health communication scholars and health professionals and has been identified as an effective strategy in promoting cancer prevention (Shi & Smith, 2016). Nevertheless, the current findings revealed that fear does not help in dissemination cancer information on social media. Future research should delve into the role of other specific emotions in the health context.

Comparing the models for the three dimensions of information diffusion, we found that the factors driving engagement are substantially different from the scale and virality of retweets. The topic is a crucial determinant of information participation; specifically, individuals are more likely to engage and discuss experience-related content. This suggests that framing the content as personal experience could be an effective strategy in involving audiences in cancer information diffusion. Messages from organizations and medical professionals are more likely to go popular and viral but are less likely to initiate discussion. One possible explanation could be that the content from organizations and medical professionals is mostly prevention or background information. To process the content usually requires people to have a high level of knowledge and thus hinders relevant discussions.

Theoretical and Practical Implications

Several insights emerge from these analytic results that can inform theoretical studies, methodology designs, and practices. First, exploring information diffusion on social media raises a different research question. Social media health campaigns are different from traditional campaigns, in which audiences are also content generators, and retweeting and commenting behaviors can change the processing and interpretation of the original content (Shi, Poorisat, & Salmon, 2018). Thus, clarifying the mechanism behind the dissemination of cancer information contributes to the literature on social media health campaigns. With this in mind, we used scale, virality, and engagement as the outcomes for online cancer information promotion, representing the different dimensions of user involvement in information diffusion on social media that can be used by future research and social media health campaigns. Second, this study examined information diffusion based on the health-specific context. Previous research seeks to investigate the driving factors of information diffusion relays on general factors, such as follower and followees, sentiment, URLs, mentions, and hashtags in the content. In this study, we not only account for the effect from these general factors but also introduced cancer-specific factors such as medical background, content topics, and fear and hope emotions. This answered the call for context-specific models of information diffusion.

The current findings thus provide strategic guidelines for the design of cancer campaigns and education on social media. In particular, to enhance the dissemination of cancer messages through social media, both the sender and content effects should be noticed. Users' credibility as indicated by account type, number of followers, number of followees, and verifying status, play a vital role in accelerating the diffusion process. To facilitate the dissemination of cancer information, health professionals may ask help from the Weibo users with a large number of followers or organizational accounts. In addition, to engage the audience in discussion, grassroots users with many followers should be a better choice.

In addition, the content factors also matter for social media cancer education programs, especially for accelerating discussions among the audience. Including hope and personal experience-related topics in the content are effective strategies for designing appealing information. For example, health professionals may provide short stories about how a cancer survivor fight with cancer to promote cancer prevention behaviors via social media, which may receive more attention and engagement among the social media users than presenting facts on cancer prevention.

Limitations and Future Work

This study has several limitations that should be kept in mind when interpreting the findings. First, only one social media platform is examined in this study. Generalizability of the findings may be limited by the function and policy of the platform; for example, keyword census and recommendation algorithms. Future research could replicate the current study in other social media platform and also in the context of other topics. Second, the whole network environment and structure, for instance, density and connectivity of the following network, may also affect the paths of the diffusion network. These factors are not included in our models and should be accounted for in future studies. Third, we used retweeting and commenting as the outcomes of cancer information diffusion. It will also be interesting in the future to look at the nature of the diffusion behavior, to differentiate between the positive and negative outcomes and whether they endorse or against the original posts and to examine their effect on the following information processing.

Conclusion

The current study proposes a multi-dimensional concept of information diffusion and applies it to analyze the diffusion of cancer-related information on Weibo. The findings indicate that sender as well as content factors influence the information diffusion. In addition, the factors driving engagement on social media are substantially different from those of scale and virality of information diffusion. This study thus contributes to the literature on information diffusion and provides practical implications on promoting cancer education on social media.

Footnote

[1] The dictionary for sentiment analysis was: https://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar (28 June 2016). This list was compiled by Hu and Liu over many years (Hu & Liu, 2004; Liu, 2012).

References

American Cancer Society. (2017, July 6). *Nearly half of China cancer deaths attributable to potentially modifiable risk factors*. ScienceDaily. Retrieved from www.sciencedaily.com/releases/2017/07/170706071915.htm

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control*. Oxford UK: Oxford university press.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on Twitter. In *Proceeding of the fourth ACM international* *conference on Web search and data mining, Hong Kong, China*. 65-74. doi:10.1145/1935826.1935845

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceeding of the The 21st international conference on World Wide Web, Lyon, France*. 519-528. doi:10.1145/2187836.2187907

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*, 192-205. doi:10.1509/jmr.10.0353

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceeding of the 43rd Hawaii International Conference on System Sciences, Honolulu, HI*. 1-10. doi:10.1109/HICSS.2010.412

Briones, R., Nan, X., Madden, K., & Waks, L. (2012). When vaccines go viral: An analysis of HPV vaccine coverage on YouTube. *Health Communication, 27*, 478-485. doi:10.1080/10410236.2011.610258

Casterline, J. B. (2001). *Diffusion processes and fertility transition: Selected perspectives*. Washington, DC: The National Academies Press.

Cha, M., Mislove, A., & Gummadi, K. P. (2009). A measurement-driven analysis of information propagation in the Flickr social network. In *Proceeding of the 18th international conference on World Wide Web, Madrid, Spain*. 721-730. doi:10.1145/1526709.1526806

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752-766.

Chaiken, S., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (Vol. 212-252, pp. 212-252). New York: Guilford Press.

Chen L., Wang X., Peng T-Q. (2018). Nature and diffusion of gynecologic cancer-related misinformation on social media: analysis of tweets. *Journal of Medical Internet Research*, *20*, e11515.

Chen, L., & Yang, X. (2018). Using EPPM to evaluate the effectiveness of fear appeal messages across different media outlets to increase the intention of breast self-examination among Chinese women. *Health Communication*. doi:10.1080/10410236.2018.1493416

Cheng, J., Adamic, L. A., Kleinberg, J. M., & Leskovec, J. (2016). Do cascades recur? In *Proceeding of the 25th International Conference on World Wide Web, Montral, Canada*. 671-681. doi:10.1145/2872427.2882993

Dodds, P. S., & Watts, D. J. (2004). Universal behavior in a generalized model of contagion. *Physical Review Letters, 92*, 218701. doi:10.1103/PhysRevLett.92.218701

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2015). The structural virality of online diffusion. *Management Science, 62*, 180-196.

Goel, S., Watts, D. J., & Goldstein, D. G. (2012). The structure of online diffusion networks. In *Proceeding of the 13th ACM Conference on Electronic Commerce, Valencia, Spain*. 623-638. doi:10.1145/2229012.2229058

Goldenberg, J., Han, S., Lehmann, D. R., & Hong, J. W. (2009). The role of hubs in the adoption process. *Journal of Marketing, 73*, 1-13. doi:10.1509/jmkg.73.2.1

Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *SIGMOD Rec., 42*, 17-28. doi:10.1145/2503792.2503797

Heo, J., Chun, M., Lee, H.W., & Woo, J-H. (2018). Social media use for cancer education at a community-based cancer center in South Korea. *Journal of Cancer Education*, *33*, 769-773.

Himelboim, I., & Han, J. Y. (2014). Cancer talk on Twitter: Community structure and information sources in breast and prostate cancer social networks. *Journal of Health Communication, 19*, 210-225. doi:10.1080/10810730.2013.811321

Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in Twitter. In *Proceeding of the 20th international conference companion on World Wide Web, Hyderabad, India*. 57-58. doi:10.1145/1963192.1963222

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics, Washington D.C., District of Columbia*. 80-88.

Kim, E., Hou, J., Han, J. Y., & Himelboim, I. (2016). Predicting retweeting behavior on breast cancer social networks: Network and content characteristics. *Journal of Health Communication, 21*, 479-486. doi:10.1080/10810730.2015.1103326

Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on Twitter. *Computers in Human Behavior, 64*, 575-583. doi:https://doi.org/10.1016/j.chb.2016.07.035

Lee, E.-J., & Tandoc, J. E. C. (2017). When news meets the audience: How audience feedback online affects news production and consumption. *Human Communication Research, 43*, 436-449. doi:10.1111/hcre.12123

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*, 1-167. doi:10.2200/S00416ED1V01Y201204HLT016

Liu, Z., Liu, L., & Li, H. (2012). Determinants of information retweeting in microblogging. *Internet Research, 22*, 443-466. doi:10.1108/10662241211250980

Loeb, S., Katz, M. S., Langford, A., Byrne, N., & Ciprut, S. (2018). Prostate cancer in social media. *Nature Reviews Urology*, *15*, 422-429.

Ma, Z., Sun, A., & Cong, G. (2013). On predicting the popularity of newly emerging hashtags in Twitter. *Journal of the American Society for Information Science and Technology, 64*, 1399-1410. doi:10.1002/asi.22844

Meng, J., Peng, W., Tan, P.-N., Liu, W., Cheng, Y., & Bae, A. (2018). Diffusion size and structural virality: The effects of message and network features on spreading health information on Twitter. *Computers in Human Behavior, 89*, 111-120. doi:10.1016/j.chb.2018.07.039

Morris, M., & Ogan, C. (1996). The Internet as mass medium. *Journal of Communication, 46*, 39-50. doi:doi:10.1111/j.1460-2466.1996.tb01460.x

O'Neill, B. (2017). Towards an improved understanding of modern health information ecology. *Social Science & Medicine, 173*, 108-109. doi:10.1016/j.socscimed.2016.12.003

Petrovic, S., Osborne, M., & Lavrenko, V. (2011). Rt to win! predicting message propagation in Twitter. *ICWSM, 11*, 586-589.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123-205): Academic Press.

Shi, J., Chen, L., Su, Y., & Chen, M. (2018). Offspring caregivers for Chinese women with breast cancer: Their social support requests and provision on social media. *Telemedicine and e-Health*. Advance online publication. doi:10.1089/tmj.2018.0176

Shi, J., Poorisat, T., & Salmon, C. T. (2018). The use of social networking sites (SNSs) in health communication campaigns: Review and recommendations. *Health Communication*, *33*, 49–56.

Shi, J., & Salmon, C. T. (2018). Identifying opinion leaders to promote organ donation on social media: Network study. *Journal of Medical Internet Research*, *20*, e7.

Shi, J., & Smith, S. W. (2016). The effects of fear appeal message repetition on perceived threat, perceived efficacy, and behavioral intention in the extended parallel process model. *Health Communication, 31*, 275-286. doi:10.1080/10410236.2014.948145

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813-838. doi:10.1037/0022-3514.48.4.813

Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). A terror management theory of social behavior: The psychological functions of self-esteem and cultural worldviews. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 24, pp. 93-159): Academic Press.

Srivastava, J., Saks, J., Weed, A. J., & Atkins, A. (2018). Engaging audiences on social media: Identifying relationships between message factors and user engagement on the American Cancer Society's Facebook page. *Telematics and Informatics*, *35*, 1832-1844.

Starbird, K., & Palen, L. (2012). (How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In *Proceeding of the ACM 2012 conference on Computer Supported Cooperative Work, Seattle, Washington, USA*. 7-16. doi:10.1145/2145204.2145212

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. doi:10.2753/MIS0742-1222290408

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? large scale analytics on factors impacting retweet in Twitter network. In *Proceeding of the 2010 IEEE second international conference on Social Computing, Minneapolis, MI*. 177-184.

Susarla, A., Oh, J.-H., & Tan, Y. (2012). Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research, 23*, 23-41. doi:10.1287/isre.1100.0339

Thomas, G. M. (2006). Building the buzz in the hive mind. *Journal of Consumer Behaviour, 4*, 64-72. doi:10.1002/cb.158

Villagran, M. (2011). Methodological diversity to reach patients along the margins, in the shadows, and on the cutting edge. *Patient Education and Counseling, 82*, 292-297. doi:10.1016/j.pec.2010.12.020

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, U.K.: Cambridge University Press.

Watson J. (2018). Social media use in cancer case. *Seminars in Oncology Nursing*, *34*, 126-131.

Watts, S. A., & Zhang, W. (2008). Capitalizing on content: Information adoption in two online communities. *Journal of the Association for Information Systems, 9*, 73-94.

Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports, 3*, 2522.

Westerman, D., Spence, P. R., & Van Der Heide, B. (2012). A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. *Computers in Human Behavior, 28*, 199-206. doi:10.1016/j.chb.2011.09.001

Wink, P., & Scott, J. (2005). Does religiousness buffer against the fear of death and dying in late adulthood? Findings from a longitudinal study. *The Journals of Gerontology: Series B, 60*, 207-214. doi:10.1093/geronb/60.4.P207

Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. *ICWSM, 10*, 355-358.

Yang, J., & Leskovec, J. (2010). Modeling information diffusion in implicit networks. In *Proceeding of the 2010 IEEE International Conference on Data Mining, Sydney, Australia*. 599-608. doi:10.1109/ICDM.2010.22

Zhang, L., & Peng, T.-Q. (2015). Breadth, depth, and speed: Diffusion of advertising messages on microblogging sites. *Internet Research, 25*, 453-470. doi:10.1108/IntR-01-2014-0021

Zhang, L., Peng, T. Q., Zhang, Y. P., Wang, X. H., & Zhu, J. J. (2014). Content or context: Which matters more in information processing on microblogging sites. *Computers in Human Behavior*, 31, 242-249.