



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Major Article

## Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging infectious disease



Lu Tang PhD <sup>a,\*</sup>, Bijie Bie PhD <sup>b</sup>, Degui Zhi PhD <sup>c</sup>

<sup>a</sup> Department of Communication, Texas A&M University, College Station, TX

<sup>b</sup> Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL

<sup>c</sup> School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX

### Key Words:

Twitter  
Crisis and emergency risk communication  
Health communication  
Social media  
Social network

**Background:** The public increasingly uses social media not only to look for information about emerging infectious diseases (EIDs), but also to share opinions, emotions, and coping strategies. Identifying the frames used in social media discussion about EIDs will allow public health agencies to assess public opinions and sentiments.

**Method:** This study examined how the public discussed measles during the measles outbreak in the United States during early 2015 that originated in Disneyland Park in Anaheim, CA, through a semantic network analysis of the content of around 1 million tweets using KH coder.

**Results:** Four frames were identified based on word frequencies and co-occurrence: news update, public health, vaccination, and political. The prominence of each individual frame changed over the course of the pre-crisis, initial, maintenance, and resolution stages of the outbreak.

**Conclusions:** This study proposed and tested a method for assessing the frames used in social media discussions about EIDs based on the creation, interpretation, and quantification of semantic networks. Public health agencies could use social media outlets, such as Twitter, to assess how the public makes sense of an EID outbreak and to create adaptive messages in communicating with the public during different stages of the crisis.

© 2018 Association for Professionals in Infection Control and Epidemiology, Inc. Published by Elsevier Inc. All rights reserved.

Emerging infectious diseases (EIDs) present novel and unfamiliar risks to the public. As a new tool for strategic communication during EID outbreaks, social media allows government agencies such as the Centers for Disease Control and Prevention (CDC) to reach wider audiences.<sup>1</sup> As a platform based on user-generated content, social media enables the public to share opinions and sentiments during these outbreaks. One such EID is measles. Measles is a highly contagious and acute illness that can lead to pneumonia, encephalitis, and death.<sup>2</sup> It was declared eliminated in the United States in 2000 as the result of the successful nationwide administration of a 2-dose vaccination (ie, measles, mumps, and rubella vaccine). However, recent years have witnessed the re-emergence of measles outbreaks.<sup>2</sup> Most of these outbreaks were associated with imported cases; at the same time, a decrease in the domestic

vaccination rate has made the country increasingly vulnerable during such outbreaks.

Presented here is a semantic network analysis of Twitter content about measles during the measles outbreak that first appeared in California during early 2015. Examining the most frequently used key words and their co-occurrences allows researchers to induce a semantic network that represents the major frames used in large amounts of text. Frames represent the cognitive structure people use in understanding and communicating issues. Through framing, media and individuals choose to highlight certain aspect of the crisis while downplaying other aspects.<sup>3</sup> This study adds to the research on crisis and emergency risk communication by demonstrating that social media users applied different frames to understand the public health crisis associated with a measles outbreak: news update frame, public health frame, vaccination frame, and political frame. In terms of research methodology, this study demonstrates the feasibility of identifying frames through semantic network analysis. Practically, the findings of the study allow public health professionals to understand how social media users make sense of an EID during different stages of the outbreak so that they can develop more effective crisis communication strategies.

\* Address correspondence to Lu Tang, PhD, Department of Communication, Texas A&M University, 4234 TAMU, College Station, TX 77843.

E-mail address: [ltang@tamu.edu](mailto:ltang@tamu.edu) (L. Tang).

Conflicts of interest: None to report.

Partially supported by a University of Alabama System Collaborative Research Initiative Grant awarded to L.T. and D.Z..

## LITERATURE REVIEW

Social media plays an essential role in the dissemination of information on EIDs. What social media users post, share, like, and comment on reflects not only what information is available, but also what they consider important. Researchers have used social media data to assess public perceptions, sentiments, and responses toward EID outbreaks such as the 2009 H1N1 outbreak,<sup>4</sup> the 2011 European *Escherichia coli* outbreak,<sup>5</sup> the 2014 Ebola virus disease outbreak,<sup>6</sup> and the 2013-2014 measles outbreak in the Netherlands.<sup>7</sup> With the exception of Lazard et al.,<sup>6</sup> all these studies examined social media contents deductively based on pre-existing categories through either manual coding or text mining based on manually coded training dataset. Semantic network of social media contents represents a new lens through which researchers could inductively investigate how the public thinks and feels during an EID outbreak without the need for a training dataset.

Semantic networks represent the semantic relationships among a set of words.<sup>8</sup> Semantic network analysis is both a theoretical framework and a quantitative text analysis method that uncovers the structure of relationships among words.<sup>9</sup> In a semantic network, word-use frequencies and co-occurrence of the most frequently occurring words represent shared meanings and common perceptions in people's minds.<sup>9</sup>

Semantic networks can be used to infer the frames used in texts. Framing is the process by which organizations and individuals choose to report or discuss an event (such as a public health crisis) by selectively highlighting certain aspects and downplaying other aspects of the event.<sup>3</sup> Researchers have explored the frames used by different stakeholders facing a crisis through the examination of semantic networks. David et al.<sup>10</sup> examined 6 pre-established frames about population issues in news articles in the Philippines by looking at weighted semantic networks between 1986 and 2000. However, a strength of semantic network analysis is it allows new and unfamiliar frames to emerge from the data inductively. For instance, Schultz et al.<sup>11</sup> studied the associative frames used by public relations professionals and news media in the United States and United Kingdom after the 2010 BP oil spill into the Gulf of Mexico and found that BP framed the oil spill as caused by external factors and downplayed internal factors (such as company's behaviors), whereas the news media adopted more complicated frames. Van der Meer et al.<sup>12</sup> compared the frames used by public relations professionals, news media, and the public in times of crises such as explosions and volcano eruptions by examining the semantic networks and found the frames used by these groups converged overtime. Tian and Stewart<sup>13</sup> compared the semantic networks based on word co-occurrence in CNN and BBC online news report about severe acute respiratory syndrome and found that although both news outlets used the public health frame, CNN used the economic impact frame, and BBC used the outbreak impact frame. Exploring frames during a measles outbreak as indicated by semantic networks by Twitter users can provide insights into how social media users make sense of the crisis and what issues concern them. Hence, the first research question (RQ) was proposed.

RQ1: What are the frames used in the Twitter discussion about the 2015 measles outbreak as indicated by semantic networks?

Crisis communication takes different forms in different stages of the crisis. Reynolds and Seeger<sup>14</sup> identified a stage-based model of crisis and emergency risk communication, which includes 5 stages: pre-crisis, initial event, maintenance, resolution, and evaluation. Each of these stages is associated with different tasks for crisis communication. Although this model was proposed as a strategic tool for guiding the crisis communication of government agencies such as the CDC, it nevertheless shows that the public goes through different

stages in information processing and sensemaking about public health crises such as a measles outbreak. As a result, they are likely to use different frames to discuss measles on Twitter during each stage of the crisis. Hence, we asked the next RQ:

RQ2: How does the use of different frames change over different stages of the outbreak?

## METHODS

### Sampling

Because the most recent measles outbreak in the United States occurred between December 2014 and April 2015,<sup>2</sup> tweets including the word *measles* posted between December 1, 2014, and April 30, 2015, were purchased from [DiscoverText.com](http://DiscoverText.com) (n = 1,154,156). Using this time frame allowed us to look at the Twitter discussion before, during, and after the outbreak.

### Data cleaning

First, raw texts were cleaned and transformed into an appropriate format to be mined.<sup>10</sup> Non-English tweets were excluded, resulting in 1,133,656 tweets. For the purpose of semantic network analysis, URLs in tweet texts were deleted. Special characters such as |, ^, and \$ as well as user names mentioned after @ were also deleted from tweet texts. Next, stop words were excluded, including conjunctions, auxiliary verbs, and transitive verbs, among others.<sup>12</sup> Plural forms of words were replaced by singular forms and high frequency words sharing the same root were combined into single words to facilitate the analysis.<sup>10</sup> Similarly, multiword phrases with the same meaning were combined.

### Semantic network analysis

Semantic network analysis was conducted to answer the RQs proposed. First, the content of tweets was analyzed to calculate the frequency of words and determine the most frequently co-occurring word pairs using KH Coder version 2.00f, a free software for analyzing text and identifying co-occurrence networks (Available for download at <https://sourceforge.net/projects/khc/>). When the data were loaded into the program, a word frequency table and a word co-occurrence network were generated. Each individual tweet was a unit of analysis and word pair co-occurrence was defined as the appearance of 2 words in the same tweet.

To explore the dynamic nature of Twitter discussion during an EID outbreak, the duration of the measles outbreak was divided into 4 stages. The pre-crisis stage (stage 1) was between December 1, 2014, and January 4, 2015. The first cases of this outbreak were reported on January 5, which marked the beginning of outbreak and thus the initial stage (stage 2).<sup>2</sup> The maintenance stage (stage 3) started on January 28, when the number of new cases started to decline and ended on March 6, when the last case in this break was reported. The resolution stage (stage 4) was between March 7 and April 17, when the CDC announced this outbreak to be officially over.<sup>2</sup> Five days were selected for each stage. To be included in the final semantic network, a word or multiword phrase must have appeared in more than 1% of tweets and be involved in the top-60 edges (connections) of each stage filtered based on Jaccard coefficient.<sup>15</sup> Jaccard coefficient<sup>16</sup> is a statistical measure widely used for assessing similarity between objects (Jaccard, 1912). The values of Jaccard coefficient vary between 0 and 1. In KH Coder, words that appear frequently in the same tweet are considered to be closely associated, and their Jaccard coefficient become closer to 1. To facilitate a more nuanced understanding of the network structure during

different stages of outbreak progression, an inductive approach was adopted to identify potential frames based on the high-frequency words and co-occurrence links among them.<sup>17</sup> The visualization of the semantic networks was accomplished using Graphviz (<http://graphviz.org>).

To explore the longitudinal changes in the use of the 4 frames identified, dictionaries containing key words associated with each frame were created based on the semantic networks identified and on the authors' further reading of tweets adopting these frames. Tweets were labeled as containing a frame based on the presence of tags (ie, frame-relevant terms). In other words, each tweet was labeled by a unique set of key words. For example, the groups of words classified as news update frame included *California, Disney, Utah, January, visitor, official, and today*. The list of words associated with the public health frame included *CDC, patient, disease, contract, infectious, and virus*, among others. The vaccine frame, which included any mention of vaccine-related issues, was associated with key terms such as *vaccine, antivaccine, unvaccinate, unvaccinated, inhaled measles vaccine, safety, immunity, and Jenny McCarthy* (an actress often associated with the anti-vaccine movement). Finally, the political frame was associated with terms such as *Obama, Republicans, Democrats, immigration, illegal, lawmaker, debate, Governor Christie, and politics*. A single tweet can be labeled as using 1 frame, multiple frames, or none of the frames. Next, percentages of tweets using different frames were calculated for the 4 stages identified and  $\chi^2$  tests were run to see whether the differences among stages were statistically significant. Bonferroni correction was conducted to account for the effects of multiple testing.

## RESULTS

Overall, original tweets ( $n = 567,391$ ) accounted for 50.05% of all tweets collected ( $n = 1,133,656$ ). As shown in Figure 1, although the numbers of tweets and retweets were similar during the most active part of the outbreak, retweets outnumbered original tweets mostly in early February during the maintenance stage.

### RQ1: Frames of the Twitter discussion of measles

Figure 2 represents the semantic network of Twitter content about measles during the entire outbreak. Four distinct frames were identified inductively based on the reading of the semantic networks and tweets containing the key words included in these semantic networks: news update frame, public health frame, vaccine frame, and political frame. The titles of these frames were coined by the authors based on the typical message carried in each frame, similar to the practice described in Odlum and Toon (2015).<sup>18</sup>

#### News update frame

The news update frame provided news and updates about suspected and confirmed cases of measles before, during, and after the outbreak. Typically, a tweet adopting the news update frame included the number of cases in a geographic location; for example, "California reports 4 more measles cases." Tweets using the news update frame did not typically contain opinions.

#### Public health frame

Tweets adopting the public health frame conveyed medical information such as symptoms of measles, methods of prevention, and treatment. For instance, the following tweet introduced a new complication of measles: "Eye complications possible with measles warn ophthalmologists." Tweets using the public health frame educated the public about measles and sometimes included behavior recommendations.

#### Vaccine frame

The vaccine frame referred to the discussion and debates about the safety and necessity of vaccination. Tweets using this frame were often emotionally charged. An example of a provaccine tweet was: "If I get measles because some nitwit talked my parents into not vaccinating me, somebody's getting their ass kicked." An example of an anti-vaccine tweet was: "Measles vaccine kills 6 kids—media blackout."

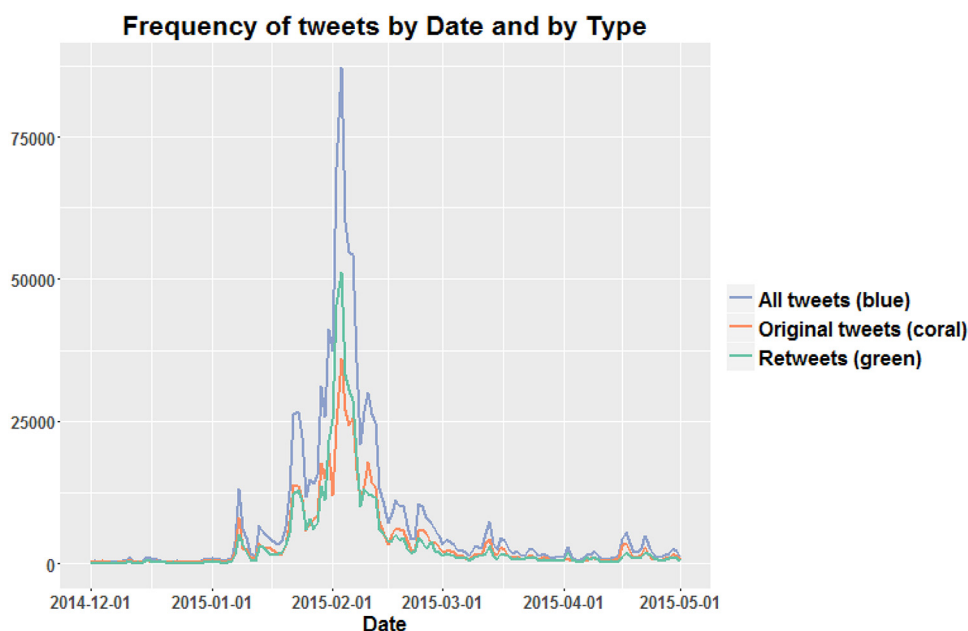


Fig 1. Number of tweets on measles by date and by type (original tweets and retweets).





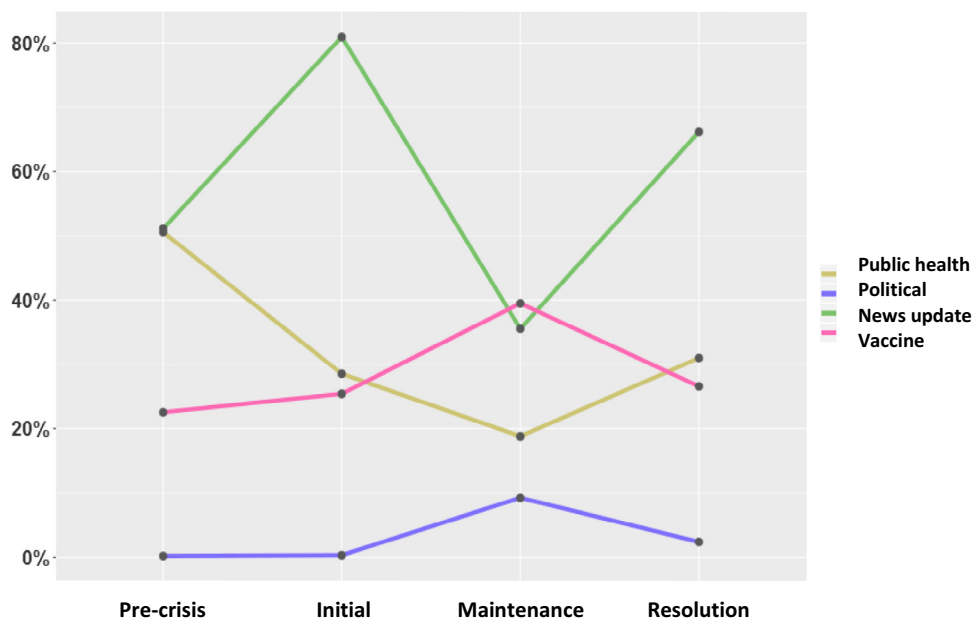


Fig 3. Percentage of tweets using 4 frames in different stages of the 2015 measles outbreak.

Table 2

Pairwise comparison of frames in measles tweets during 4 stages

	Pre-crisis vs initial		Initial vs maintenance		Maintenance vs resolution	
	P value	Adjusted P value	P value	Adjusted P value	P value	Adjusted P value
Updates	.00*	.00*	.00*	.00*	.00*	.00*
Health	.00*	.00*	.00*	.00*	.00*	.00*
Vaccine	.0003*	.0006*	.00*	.00*	.00*	.00*
Political	.34	.34	.00*	.00*	.00*	.00*

NOTE: The degree of freedom of all  $\chi^2$  tests is 1.

\*Adjusted P value < .001.

adjusted  $P < .001$  level, except for the use of the political frame in pre-crisis and initial stages.

## DISCUSSION

Social media allows the assessment of public opinions, sentiments, and responses during an EID outbreak. Our study examines the frames used in Twitter discussion during the 2015 measles outbreak through a semantic network analysis.

This study finds that around half of tweets are retweets. Furthermore, retweets outnumbered original tweets during the early days of the maintenance stage. This stands in contrast with the findings of previous studies. For instance, Liu, Kliman-Silver, and Mislove<sup>19</sup> found that overall around 30% of tweets are retweets. Radzikowski et al<sup>20</sup> studied the tweets about vaccines during the later stages of the 2015 measles outbreak (the maintenance stage in our study) and found around 44% of tweets in their data corpus were retweets. Our data also suggest the highest rate of retweets during this stage, although our data were about measles instead of vaccine. It is possible that the maintenance stage sees more reflection and activism when the immediate threat of the outbreak has been contained and thus is associated with more retweets.

The current study identified 4 major frames that emerged organically from the semantic network based on word frequencies and co-occurrence. Each frame highlighted an important dimension of

measles-related discussion on Twitter. Furthermore, different stages of the outbreak witnessed fluctuations in each frame's popularity. During the pre-crisis stage, the news update frame and public health frame were the 2 most dominant frames used, whereas the vaccine frame was rarely used and the political frame was almost never used. During the initial stage, use of the news update and vaccine frames increased, although the use of the public health frame actually went down. During the maintenance stage, the vaccine frame became the most frequently used frame, followed by the news update frame, public health frame, and political frame. As the outbreak drew to an end, the use of frames revert to the pattern observed in the pre-crisis stage, with news update being the most used frame, followed by the public health frame, vaccine frame, and political frame.

There are 2 possible explanations for the changes in how measles is framed on Twitter. It is possible that the frames on Twitter are influenced by the frames set by the mainstream media. For instance, Lee and Basnyat<sup>21</sup> studied media framing of the 2009 H1N1 outbreak in Singapore and suggested that although traditional media predominantly used the update frame and prevention frame throughout the break, their use of frames diversified to some extent in the later stages of the crisis to include more personal frames and social frames. Our semantic network shows that traditional news media (eg, CNN, CBS, Reuters, and AP) featured prominently in the discussion about the measles outbreak. Future research could examine the intermedia agenda setting process between traditional news media and social media in covering and framing public health crises associated with EIDs to see if Twitter frames are influenced by the frames used in the traditional news media or if the public may influence the frames used by traditional news media through voicing their concerns and opinions in Twitter. The other possible explanation for the changes in the use of frames is psychological. According to the psychometric paradigm, people's response to a risk is decided by the dread, catastrophic potential, controllability, and familiarity of said risk.<sup>22</sup> It is possible that during the early stages of a measles outbreak, the public is unsure about the seriousness of the risk and the controllability of the outbreak and therefore prefers the update frame and the public health frame. These 2 frames provide information that can help the public assess the scale and severity of the crisis. However, as the outbreak

develops, people perceive the risks associated with measles as more familiar and more controllable. As a result, they start to try to make sense of the crisis by deciding whom they should blame for the outbreak, thereby leading to the increased use of the vaccine frame, which blames parents who refuse to vaccinate their children, as well as the political frame, which blames illegal immigrants or the government for the outbreak. Through surveys, future research could establish the relationship between social media users' risk perception and the frames they apply in discussing EIDs.

The high frequencies of the political frame during the maintenance stage, some of which state that illegal immigrants were responsible for the measles outbreak, also indicates the spread of fake news on Twitter. Recently, Vosoughi, Roy, and Aral<sup>23</sup> found that false news stories spread further, faster, deeper and more broadly on Twitter than truthful stories. Compared with truth, false news stories only took one-sixth of the time to reach 1,500 people and were 70% more likely to be retweeted. These findings have particularly important implications for health care professionals during infectious disease emergencies. When rumor or misinformation gains high virality, informing and educating the public would become very challenging and difficult.

Radzikowski et al. studied the measles vaccination narrative in Twitter by examining the semantic networks of popular hashtags and retweeting patterns in the aftermath of the outbreak. They found that although official public health agencies such as the CDC and World Health Organization have all entered the social media arena, mainstream media coverage about key health issues still have the power to lead online public participation. Our study confirmed some of the findings of Radzikowski et al.<sup>20</sup> that mainstream news coverage of health topics actually played a very important role in leading online audience attention and shaping public debate on social media and political frame was the most observed (9.3%) during the maintenance stage when the vaccine frame also came to its peak (39.5%). Our study also extended Radzikowski et al.<sup>20</sup> by extending the observation time, looking at Twitter discussion of this outbreak as a whole from a risk communication perspective, and comparing the content patterns in different stages of the outbreak.

This study proposes and tests a method for assessing the frames used in social media discussions about EIDs based on the creation, interpretation, and quantification of semantic networks. This method can be used to study the public response to different EID outbreaks as well as other ongoing public health crises, such as heart disease and obesity. Instead of looking for known frames used in social media, this method allows the identification of unique frames associated with different public health crises.

In terms of this study's practical implications, our study shows that agencies could use social media content on platforms such as Twitter to assess how the public makes sense of an EID outbreak and to create adaptive messages in communicating with the public during different stages of the crisis. For instance, if it is detected that the public tends to use the vaccine frame during a certain stage of the crisis, public health agencies could design and disseminate specific social media messages to spread useful information and combat misconceptions.

This study only represented an early-stage effort at mining the contents about EIDs on social media. Future studies can take a number of directions. It is possible that individuals, government agencies, nonprofit organizations, and news media might use different frames in tweeting about measles before, during, and after

an EID outbreak. Further research should compare these different stakeholders' tweets. In addition, it would be of interest to study the retweeting network about EIDs (as done in Radzikowski et al, 2016). Studying the structure of retweeting networks can shed lights on how information and opinion is diffused among Twitter users. It would be useful to compare the retweeting networks during different stages of the outbreak to see if information flows differently during these stages and explore the roles played different stakeholders (eg, traditional news media, nonprofit organizations, and corporations) during across the stages.

## References

- Gesser-Edelsburg A, Stoloro N, Mordini E, Billingsley M, James JJ, Green MS. Emerging infectious disease (EID) communication during the 2009 H1N1 influenza outbreak: literature review (2009-2013) of the methodology used for EID communication analysis. *Disaster Med Public Health Prep* 2015;9:199-206.
- Clemmons NS, Gastanaduy PA, Fiebelkorn AP, Redd SB, Wallace GS. Measles – United States, January 4–April 2, 2015. *MMWR Morb Mortal Wkly Rep* 2015;64:373-6.
- Scheufele DA. Framing as a theory of media effects. *J Commun* 1999;49:103-22.
- Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE* 2010;5.
- Gaspar R, Gorjão S, Seibt B, Lima L, Barnett J, Moss A, et al. Tweeting during food crises: A psychosocial analysis of threat coping expressions in Spain during the 2011 European EHEC outbreak. *Int J Hum Comput Stud* 2014;72:239-54.
- Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *Am J Infect Control* 2015;43:1109-11.
- Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. *J Med Internet Res* 2015;17:e128.
- Monge PR, Contractor NS. *Theories of communication networks*. New York: Oxford University Press; 2003.
- Doerfel M. What constitutes semantic network analysis? A comparison of research and methodologies. *Connect* 1998;21:16-26.
- David C, Legara E, Atun J, Monterola C. News frames of the population issue in the Philippines. *Int J Commun* 2014;8:21.
- Schultz F, Kleinnijenhuis J, Oegema D, Utz S, van Atteveldt W. Strategic framing in the BP crisis: A semantic network analysis of associative frames. *Pub Relat Rev* 2012;38:97-107.
- Van der Meer TGLA, Verhoeven P, Beentjes H, Vliegenthart R. When frames align: the interplay between PR, news media, and the public in times of crisis. *Public Relat Rev* 2014;40:751-61.
- Tian Y, Stewart CM. Framing the SARS crisis: a computer-assisted text analysis of CNN and BBC online news reports of SARS. *Asian J Commun* 2005;15:289-301.
- Reynolds B, Seeger MW. Crisis and emergency risk communication as an integrative model. *J Health Commun* 2005;10:43-55.
- Danowski JA. Network analysis of message content. In: Richards WD Jr, Barnett GA, editors. *Progress in Communication Sciences*, Vol. 12. Norwood, NJ: Ablex; 1993. p. 197-221.
- Jaccard P. The distribution of the flora in the alpine zone. *New Phytologist* 1912;11:37-50.
- Rule A, Cointet J-P, Bearman PS. Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790-2014. *Proc Natl Acad Sci U S A* 2015;112:10837-44.
- Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *Am J Infect Control* 2015;43:563-571.
- Liu Y, Kliman-Silver C, Mislove A. The tweets they are a-changin': Evolution of Twitter users and behavior. AAAI. 2014. p. 305-14. Proceedings of the 8th International Conference on Weblogs and Social Media.
- Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Pub Health Surveill* 2016;2(1).
- Lee ST, Basnyat I. From press release to news: mapping the framing of the 2009 H1N1 influenza pandemic. *Health Commun* 2013;28:119-32.
- Slovic P. *The perception of risk*. London: Earthscan; 2000.
- Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science* 2018;359:1146-51.