Heliyon 8 (2022) e09994

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Understanding COVID-19 response by twitter users: A text analysis approach

Digvijay Pandey [a,*], Bandinee Pradhan [b], Wangmo [c]

[a] Department of Technical Education, IET, Dr. A.P.J. Abdul Kalam Technical University Uttar Pradesh, Lucknow 226021, India
[b] PDPU, Gandhinagar, India
[c] Sherubtse College, Royal University of Bhutan, Bhutan

ARTICLE INFO

Keywords:
COVID-19
Twitter
Public health
Public opinion
Sentiments

ABSTRACT

COVID-19 outbreak has caused a high number of casualties and is an unprecedented public health emergency. Twitter has emerged as a major platform for public interactions, giving opportunity to researchers for understanding public response to the outbreak. The researchers analyzed 100,000 tweets with hashtags #coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #COVID19, #COVID-19, #epitwitter, #ihavecorona, #StayHomeStaySafe, #TestTraceIsolate. Programming languages such as Python, Google NLP, and NVivo are used for sentiment analysis and thematic analysis. The result showed 29.61% tweets were attached to positive sentiments, 29.49% mixed sentiments, 23.23 % neutral sentiments and 18.069% negative sentiments. Popular keywords include "cases", "home", "people" and "help". We identified "30" such topics and categorized them into "three" themes: Public Health, COVID-19 around the world and Number of Cases/Death. This study shows twitter data and NLP approach can be utilized for studies related to public discussion and sentiments during the COVID-19 outbreak. Real time analysis can help reduce the false messages and increase the efficiency in proving the right guidelines for people.

1. Introduction

On December 31, 2019, an unknown pneumonia-like disease was detected in Wuhan, China and was reported to WHO country-office in China. In Early January, it was declared as an International public health emergency. In February, WHO announced the name of the new coronavirus as COVID-19. Since then, people all around the world are fighting to find a vaccine for the virus and the authorities are taking measures to keep the public safe from the virus. These measures include easy and efficient access to testing and results, rigorous contact tracing, consistent science-based messaging, quarantines and a genuine commitment to clamping down on socializing. The recent COVID-19 or Coronavirus pandemic is one such topic that has been trending on twitter. Ever since the outbreak in China, the global situation is worsening. As of April 28, 2020, globally, there were 2,954,222 cases of COVID-19. The cases are increasing globally with 960,916, highest cases alone in the United States. The virus is primarily spread between people during close contact, often via small droplets produced by coughing, sneezing, or talking. To stop the spreading of the virus, authorities worldwide were implementing travel bans, lockdowns, quarantines, curfews, stay-at-home orders, sanitizations, and public facilities closures.

New technologies such as Twitter, Facebook and Redditt have played an important role in allowing people to express their opinion on social media platforms. These technologies have made it possible for users to create, modify, and share information (Kietzmann et al., 2011). Moreover, Kumar, 2020 asserts that this has brought the power of big data to researchers to carry out sentiment analysis, explore key factors from a specific community. Taking the current situation, WHO has provided guidelines to the general public and authorities on handling the pandemic. Therefore, this research will help provide information about factors related to COVID-19 and extend the current methodology in this new context.

2. Using twitter to measure Public Opinion

With 330 million monthly active users known in 2020, Twitter is one of the perfect mediums for the people to disseminate information. Social networking has been a representation of the arrival of emerging technology for modern technological environments. The web-based media

* Corresponding author.
E-mail address: digit11011989@gmail.com (D. Pandey).

https://doi.org/10.1016/j.heliyon.2022.e09994
Received 12 January 2021; Received in revised form 11 November 2021; Accepted 13 July 2022

became an impact of the event's latest advances. Studies that depend upon the Social Mediated Crisis Communication (SMCC) model have revealed that folks depend upon social and traditional media for data during emergencies (Austin et al., 2012; Liu et al., 2013). As a feature of a bigger media biological system, informal communication destinations became a facet of the range of authority utilized by associations to succeed in and include people during emergencies (Hanna et al., 2011; Keim and Noji, 2011). Studies have underlined that the determination of media (customary versus social) matters as associations endeavor to illuminate general society around public emergencies (Austin et al., 2012; Jin et al., 2014; Keim and Noji, 2011; Liu et al., 2011, 2015; Schultz et al., 2011). The contents shared on social media sites provide information as evidence during times of public health threats. This is often particularly significant for things analyzed in light of COVID-19 global pandemic as Twitter has become a replacement medium used to circulate recent information on COVID-19 by the health experts, journalists, government authorities and other genuine sources around the world.

Specifically, understanding of the profound, rich, and altered correspondence about COVID-19 on Twitter could encourage the continuous responses of people with medical concerns and the right information about it. The government and health agencies need to provide information to people regarding the potential danger as their community or country transforms into a red zone with numerous cases of coronavirus infection. Giving the open data about how an infection of COVID-19 is spread and what government pioneers do to forestall forthcoming spread may stop alarm and deceived correspondence. With more than 27 million deaths recorded thus far, the government need to reassure its citizens' safety and secure the overall population's health.

Twitter binds everybody for the overall cause during emergencies. The swine influenza outbreak in the US in 2009 drew a lot of attention from Twitterati. So, researchers studied the primary evidence-based scientific research by applying a qualitative methodology (Braun and Clarke, 2006) to tweets generated during the peak of the swine influenza epidemic. Moreover, during the 2010 seismic tremor, the American Red Cross had disentangled its raising money procedure and utilized Twitter as a discussion whereby people could tweet five-digit numbers and promptly give to the Haiti calamity subsidize (Manjoo, 2010). Increments in tweets pertaining to the catastrophe through posts/retweets containing “Haiti”' or “'seismic tremor”' were seen after the calamity, featuring that associations were conveying about the actual occasion.

According to the Centers for Disease Control and Prevention (CDCP, 2016a), in 2014 the deadliest outbreak of Ebola epidemic in history occurred with about 22,000 cases. In response to the rising Ebola concern among the general public, the CDC commenced a live Twitter conversation with the general public in an effort to provide accurate information regarding the transmission of the disease. Another evidence-based research has examined the utility of Twitter for gaining insight into communicable disease outbreaks (Chew and Eysenbach, 2010; Signorini et al., 2011; Kostkova et al., 2014) and Ebola (Oluwafemi et al., 2014). The research offered unique insight into data driven qualitative tweets associated with communicable disease outbreaks. Another research (Diddi and Lundy, 2017) explored how four different health-related organizations used their Twitter accounts to speak about varied aspects of carcinoma during the month of October, which is observed as Breast Cancer Awareness month. All the tweets by these associations were analyzed for the presence of the theoretical parameters of the Health Belief Model (HBM) and in this way the examination exhibited while various associations shared important breast cancer related content on Twitter, each utilized the online media stage with an alternate style, clear evidence through specialization in different types of HBM constructs while publishing breast cancer-related tweets.

As proposed by SMCC, associations must consider the three sorts of people to interact with their online media messages so on to enhance the adequacy of imparting through web-based media (Austin et al., 2012; Fraustino et al., 2012; Jin et al., 2014; Liu et al., 2013). Utilizing retweets could get online media makers to spread data to supporters. Moreover, with the presentation of online media, associations have taken on a human-like quality, rather than stay an inaccessible element (Notter and Grant, 2012). As the worldwide community faces the coronavirus pandemic together, Twitter assists individuals to find reliable data, connect with others, and follow what is going on progressively. There is a growing research enthusiasm for analyzing tweets according to the thought they express. This interest may be a result of the enormous proportion of messages that are posted regularly in Twitter which contain significant data for the receptive mind. Such an ordinary correspondence system would undoubtedly get the opportunity to arrange the extending level of instinct required by the segment of people for the most part. It encourages the users to be propelled by stories of courageous acts, positive examples, and global efforts to fight against the pandemic. On the contrary side, Twitter is effectively combating to prevent the deception or harm which will come to users from posts on their platform.

3. Research methodology

A five-step process was followed to collect, process, analyze and derive insights/inferences from COVID-19 tweets as shown in the flow chart below:

In Figure 1 the steps 1 to 3 were performed using custom written Python code on Anaconda (Spyder IDE) and step 4 was done on Python, Microsoft Excel 2016, Tableau Public Desktop 10.3 (Sarlan et al., 2014).

Step 1. Data collection

The scope of the analysis was to examine the public expressions, sentiments and interest/focus areas when there was a sudden spike in COVID-19 cases along with lockdown in most parts of the world, hence the time period was fixed as 12-March-2020 to 15-Apr-2020. The data (Twitter day-wise tweets data) was manually downloaded from a publicly accessible data science platform Kaggle.com using two URLs post logging in by the analyst's profile on the website. The data had only those tweets (original tweets, no retweets) that contained at least one of the following hashtags – #coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #COVID-19, #COVID_19, #epitwitter, #I havecorona, #StayHomeStaySafe, #TestTraceIsolate

Step 2. Data collation & preparation

Once the 35 CSV files were placed in a folder, the next task was to combine the data from multiple files and create a master dataset. While combining the files the following were considerations were taken:

A. Since the NLP sentiment tagging services are not available for all languages and for the sake of uniformity, only English tweets were filtered from the dataset using the ‘lang’ column in the dataset which had value ‘en’ for English.
B. Owing to data processing constraints as well as for the sake of brevity: 100K tweets were sampled from the consolidated dataset in English from #A above using simple random sampling accomplished by pandas. Sample function in Python.

A python code was written to accomplish this. The compiled data was observed to have 20 million tweets in 66 languages which was filtered on language English and sampled (using simple random sampling) to get 100K tweets[1] which were saved in a CSV file.

4. Illustration of data filtering for final (analysis) dataset

Data cleaning was not used in the preparation step so the hyperlinks and hashtags were retained in the ‘text’ (tweet content) column on the final dataset. This was done to preserve the original nature of the tweet

[1] 100.30K tweets were filtered for the analysis (final dataset) where 30 random tweet records were used in code testing and dropped from further stages of the analysis.

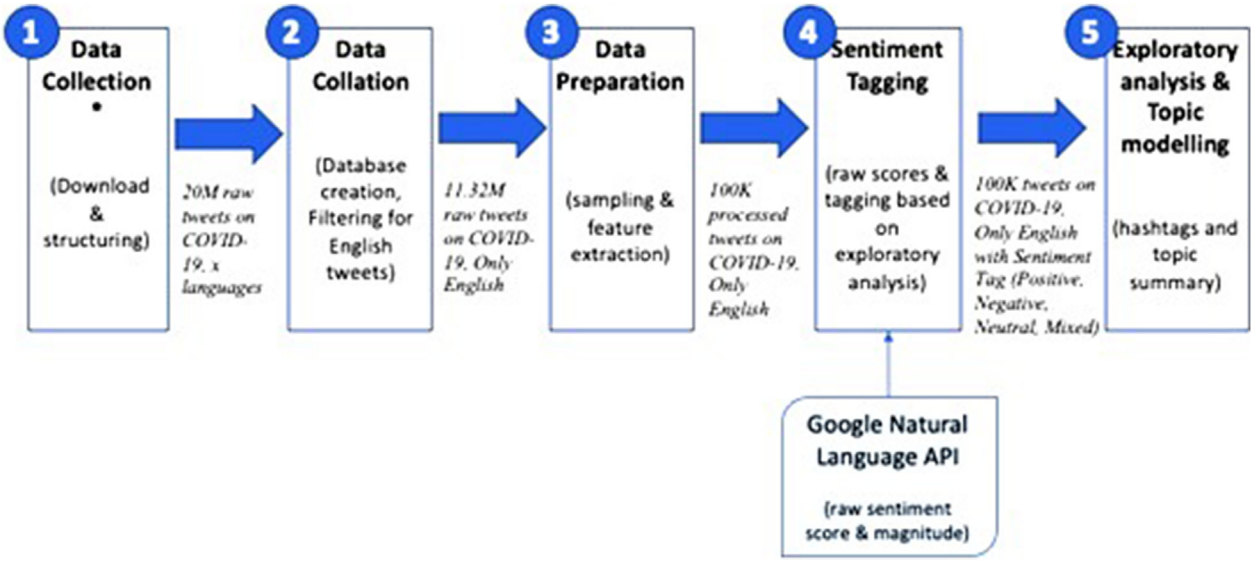

Figure 1. Sentiment Analysis Process. Source: Created by authors.

which would help in contextual sentiment tagging of the tweet. Several industry analysts such as () do cleaning of hyperlinks, hashtags, word stemming and lemmatization, stop word and special characters removal at this step-in social media text analysis; however, this was not done in the current analysis due to the aforementioned reason.

5. Sentiment tag for each tweets

Once the final dataset was ready in the CSV format; the next step was to tag the sentiment score and magnitude. Google Cloud Natural language was used for this purpose along with Python code. The algorithm iterated over each tweet in the 100K final dataset where it called the Google NLP sentiment tagging API to get the sentiment score and sentiment magnitude. The score defined the polarity of the sentiment (−1 being most negative to +1 being most positive) whereas magnitude (0 to infinity) defined the strength of the emotion.

Post tagging the sentiment score and magnitude on each tweet, threshold for tagging each sentiment was selected. Per Google NLP API documentation this is subjective to the context of analysis being done using Python code. So, the following was used to tag each tweet into one of the four sentiments – positive, negative, mixed and neutral as per Table 1 given below:

'The score, magnitude and sentiment tag were added as a new column in the dataset which was exported to CSV format to be used as a compatible file for insight generation and analysis via Python, Excel and Tableau. This file was called 'the tagged dataset'.

6. Results

A qualitative approach was used to further develop themes in the twitter data. The researchers have used NVIVO12 to carry out the analysis. Specifically, Braun and Clarke (2006) six step process was used: (1) Becoming familiar with the data, (2) Generate initial codes, (3) Search for themes, (4) Review themes, (5) Define themes, and (6) Write-up. Finally, the researchers developed themes related to COVID-19 and the popular tweeted words included "COVID-19", "people", "cases", "new", "home", help" and "health". See Table 2 and Figure 2. The result is presented in Figure 3.

Considering the enormous number of words that can be seen in Table 2, it is notable how well this word significantly influenced the

Table 1. Sentiments by tweets.

Score	Magnitude	Sentiment Tag
≥0.2	>0.25	Positive
<=-0.2	>0.25	Negative
-0.2 to +0.2	>0.25	Mixed
Any (-1 to 1)	≤0.25	Neutral

Table 2. TOP 30 Popular words.

Word	Count	Weighted Percentage
#COVID-19	49082	2.47%
#coronavirus	48538	2.44%
#COVID	12716	0.64%
amp	11188	0.56%
people	10059	0.51%
#coronaviruspandemic	6561	0.33%
cases	6384	0.32%
new	5659	0.28%
just	5464	0.27%
coronavirus	5302	0.27%
home	4986	0.25%
get	4941	0.25%
time	4920	0.25%
help	4794	0.24%
like	4579	0.23%
pandemic	4539	0.23%
need	4373	0.22%
COVID	4338	0.22%
one	4230	0.21%
#coronavirusoutbreak	4205	0.21%
stay	4122	0.21%
health	4068	0.20%
world	3910	0.20%
virus	3789	0.19%
please	3602	0.18%
today	3423	0.17%
day	3362	0.17%
support	3154	0.16%
know	3126	0.16%
deaths	3093	0.16%

Source: Created by authors.

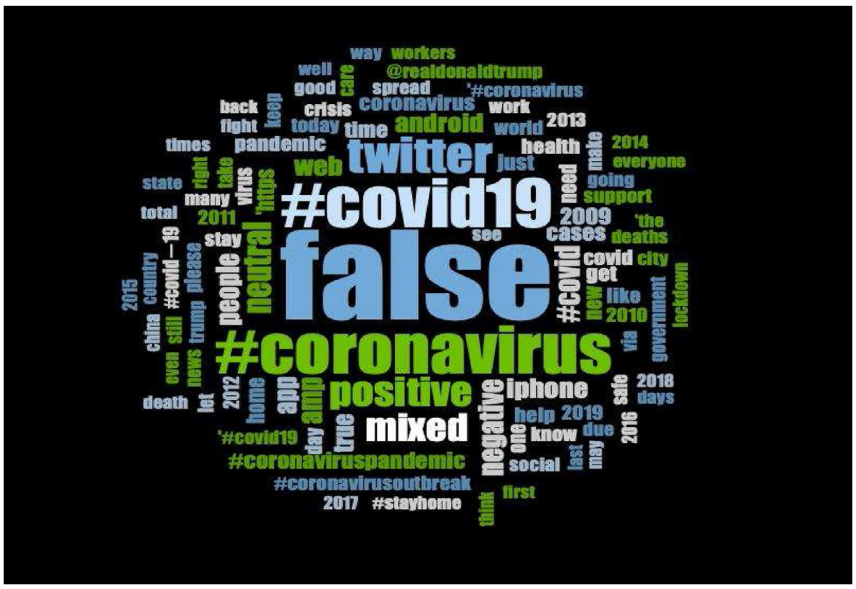

Figure 2. The word cloud of the most popular keyword. Source: Created by authorsCOVID-19 related themes.

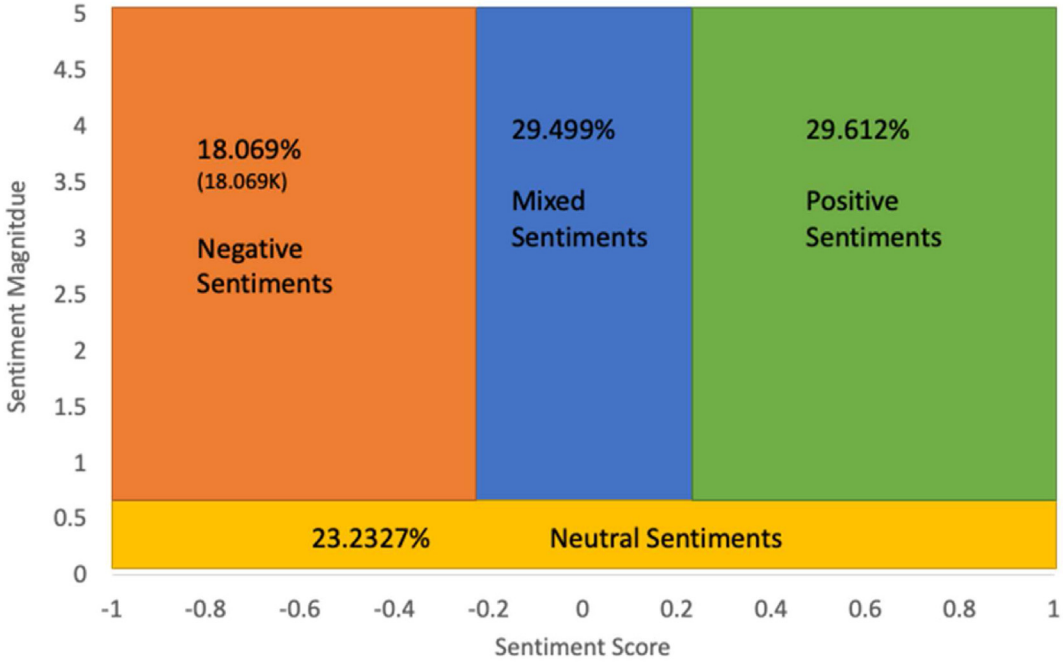

Figure 3. Sentiments by tweets.

contemporary COVID-19 pandemic narrative. During the COVID-19 outbreak, words such as "Corona" had a high number of tweets.

The ideas have been determined by individually picking relevant topics and employing the text search query in NVIVO 12, which has been utilised during the content analysis method. Table 3 reveals a high prevalence of tweets with the topics face mask, quarantine, self-quarantine, Covid Test, Lockdown, and so forth. Some of the more important words used throughout these issues included mask, face shield, home quarantine, self-quarantine, test kits, rapid test, safety, and others. Even though it's possible that such an issue seems tied to something like a separate group of circumstances, it's important to note that almost all of the tweets analysed contain a word connected with COVID-19 as well as its variants.

7. Discussion

The general discourse and emotions of COVID-19-related Twitter messages were examined in this research. Between March 12 and April 15, 2020, individuals explored the top important themes shown in Table 3. The COVID-19 subjects are better understood owing to the topic modelling. Numerous conclusions can be drawn from the findings. To start with, there are a number of references to "quarantine" throughout the discussion, meaning that people value it higher than other issues. The second most discussed issue among individuals is "lockdown." Third, respondents were worried about "safety," as well as communications circulating all around them. This indicated that respondents are saying "stay home, stay safe" increasingly frequently. Fourth, notwithstanding

Table 3. List of theme, topic, related words, and number of references on twitter.

Theme	Topic	Related words	Number of references
Public Health	Facemask	mask, facemask, face shield, PPE	5114 references, 51.14 % coverage
	Quarantine	Home quarantine, self-quarantine	12105 references, 128.05 % coverage
	COVID-19 Test	Test kits, rapid test	2944 references, 0.02 % coverage
	Lockdown	COVID-19 lockdown	9470 references, 0.08 % coverage
	Social Distancing	1 feet distance	4346 references, 0.05 % coverage
	Safety	Stay home, stay safe	8567 references, 0.05% coverage
	COVID-19 Vaccine	Vaccine	1204 references, 0.01 % coverage
COVID-19 cases around the world	COVID-19 in the United States	Lockdown in the US	5259 references, 0.05 % coverage
	UK	UK lockdown, Immunity	3256 references, 0.01 % coverage
	Italy	Italy lockdown	2623references, 0.02 % coverage
	Wuhan, China	Criticism from media	4777 references, 0.04 % coverage
COVID-19: No of new cases and Death	New cases	New cases, confirmed cases, active cases	6335 references, 0.05 % coverage
	Death rate	Death poll, COVID-19 death	3654 references, 0.03 % coverage

their fears, a majority of participants expressed optimistic and mixed sentiments, indicating that many people predict an increase in the incidence of instances. Finally, vaccinations have been the least reported topic during the study period, indicating that the disease is still very much in the initial stages and that just a few discussions about it when the vaccine would've been ready were taking place.

8. Limitations and future research

The researchers used only sample trending hashtags to collect the data. As the situation is ongoing and evolving new trends would be coming up. For example, COVID-19 vaccine was not a popular trend during the research period. Another limitation is “global sample”, since the tweets can be taken as a global representation and is indicative of the twitter “users” opinion. However, twitter does provide real time data to conduct analysis which can be valuable. Another limitation is the language, as the researchers have only analyzed the tweets in English language. These limitation gaps can be bridged in future research. Future research can delve deeper into the themes emerged in this study. For example, “lockdown”, “safety” and “vaccines”. Another avenue is to look into other trends such as “misinformation”, “politics in the time of COVID-19”.

9. Conclusion

This proposed research, particularly attempting to analyse people's emotions and feelings throughout a COVID-19 outbreak, has been an accomplishment. Throughout this study, the Twitter posting interface was chosen to ensure the reliability of the findings as well as the ease of accessing individual Twitter posts. This study demonstrates how Twitter data may be used to measure public sentiment amid emergencies like COVID-19. Even throughout the study, it was revealed that nearly all states were tweeting about COVID-19 with positive views, showing that all of those people had already become accustomed to COVID-19 and, as a result, the survival rate had increased over time. The results of this analysis revealed that subjects such as "preventive methods to combat COVID-19," "public health," and "COVID-19 cases and mortality rate" were frequently discussed. The sentiment analysis suggested maximum users showed “positive” and “mixed” emotions. This type of analysis can be helpful for government and healthcare authorities to understand and react to public emergencies. It can also be utilized to ensure trust in the public.

Declarations

Author contribution statement

Bandinee Pradhan, Digvijay Pandey, Wangmo, GadeeGowwrii: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Austin, L., Fisher Liu, B., Jin, Y., 2012. How audiences seek out crisis information: exploring the social-mediated crisis communication model. J. Appl. Commun. Res. 40 (2), 188–207.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology, Qualitative. Res. Psychol. 3 (2), 77–101.

Chew, C., Eysenbach, G., 2010. Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 5 (11).

Diddi, P., Lundy, L.K., 2017. Organizational Twitter use: content analysis of tweets during breast cancer awareness month. J. Health Commun. 22 (3), 243–253.

Fraustino, J.D., Liu, B., Jin, Y., 2012. Social media use during disasters: a review of the knowledge base and gaps. In: National Consortium for the Study of Terrorism and Responses to Terrorism. http://www.cridlac.org/digitalizacion/pdf/eng/doc19270/doc19270.htm.

Hanna, R., Rohm, A., Crittenden, V., 2011. We’re all connected: the power of the social media ecosystem. Bus. Horiz. 54 (3), 265–273.

Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C.T., Ramakrishnan, N., 2014. Misinformation propagation in the age of twitter. Computer 47 (12), 90–94.

Keim, M.E., Noji, E., 2011. Emergent use of social media: a new age of opportunity for disaster resilience. Am. J. Disaster Med. 6 (1), 47–54.

Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. Social media? Get serious! Understanding the functional building blocks of social media. Bus. Horiz. 54 (3), 241–251.

Kostkova, P., Szomzsor, M., Louis, C.S., 2014. #swineflu: the use of twitter as an early warning and risk communication tool in the 2009 Swine flu pandemic. ACM Transact. Manage. Inform. Syst. 5 (8).

Kumar, A., 2020. COVID- Tweets Collation and Sentiment Tag with Google API [Source Code], Version 4.0. https://github.com/AshwiniKmr1/COVID-_tweets_text_analytics/blob/main/COVID-19-tweets-collation-google-api-sentiment-tag_v4.py.

Liu, B.F., Austin, L.L., Jin, Y., 2011. How publics respond to crisis communication strategies: the interplay of information form and source. J. Publ. Relat. Res. 37, 345–353.

Liu, B.F., Jin, Y., Austin, L.L., 2013. The tendency to tell: understanding publics' communicative responses to crisis information form and source. J. Publ. Relat. Res. 25 (1), 51–67.

Liu, B., Fraustino, Jin, Y., 2015. How disaster information form, source, type, and prior disaster exposure affect public outcomes: jumping on the social media bandwagon? J. Appl. Commun. Res. 43 (1).

Manjoo, F., 2010. How Black People Use Twitter: the Latest Research on Race and Microblogging. http://slate.com/articles/technology/technology/2010/08/how_black_people_use_twitter.

Notter, J., Grant, M., 2012. Humanize: How People-Centric Organizations Succeed in a Social World. Que publishing.

Oluwafemi, S., Gabarron, E., Wynn, R., 2014. Ebola, Twitter, and misinformation: a dangerous combination? BMJ Clin. Res. g6178.

Sarlan, A., Nadam, C., Basri, S., 2014. Twitter sentiment analysis. In: Proceedings of the 6th International Conference on Information Technology and Multimedia. IEEE, pp. 212–216.

Schultz, F., Utz, S., Göritz, A., 2011. Is the medium the message? Perceptions of and reactions to crisis communication via twitter, blogs and traditional media. J. Publ. Relat. Res. 37 (1), 20–27.

Signorini, A., Segre, A.M., Polgreen, P.M., 2011. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 6 (5), e19467.