



Identification of affective valence of Twitter generated sentiments during the COVID-19 outbreak

Ruchi Mittal¹ · Amit Mittal² · Ishan Aggarwal³

Received: 13 October 2020 / Revised: 16 July 2021 / Accepted: 17 July 2021 / Published online: 27 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

This study aims to conduct text mining of affective valence of the sentiments generated on social media during the COVID-19 and measure their association with different outcomes of the disease. 50,000 tweets per day over 23 days during the pandemic were extracted using the VADER sentiment analysis tool. Overall, tweets could effectively be classified in terms of polarity, i.e., “positive,” “negative” and “neutral” sentiments. Furthermore, on a day-to-day basis, the study identified a positive and significant relationship between COVID-19-related (a) global infections and negative tweets, (b) global deaths and negative tweets, (c) recoveries and negative tweets, and (d) recoveries and positive tweets. No significant association could be found between (e) infections and positive tweets and (f) deaths and positive tweets. Furthermore, the statistical analysis also indicated that the daily distribution of tweets based on polarity generates three distinct and significantly different numbers of tweets per category, i.e., positive, negative and neutral. As per the results generated through sentiment analysis of tweets in this study, the emergence of “positive” tweets in such a gloomy pandemic scenario shows the inherent resilience of humans. The significant association between news of COVID-19 recoveries and positive tweets seems to hint at a more optimistic scenario whenever the pandemic finally comes to an end or is controlled. Such public reactions—for good—have the potential to go viral and influence several others, especially those who are classified as “neutral” or fence-sitters.

Keywords Coronavirus · COVID-19 · SARS-CoV-2 · Pandemic · Affective valence · Twitter · Sentiment analysis

1 Introduction

The world is witnessing an explosion of social networks (SNs) such as Twitter where people exchange information, opinions and ideas (Ombabi et al. 2020). The huge volume of data generated on such SNs can be processed to derive meaningful information using a computational technique known as natural language processing (NLP). Sentiment

analysis (SA) is a popular technique of NLP. SA based textual data aim to use text mining, linguistics and statistical knowledge techniques to automatically assign predefined sentiment labels (affective valence) to the text generated online (Alowaidi et al. 2017; Ombabi et al. 2020). In a study based on text mining of tweets generated during the 2009 H1N1 disease outbreak, it was discovered that the outbreak could have been predicted at least a week in advance based on information spreading on Twitter (Szomszor et al. 2011). There was an eerie prediction of a new Coronavirus where it was mentioned about horseshoe bats being a reservoir of Coronaviruses and civets potential as amplifiers of the disease. In the context of the recent outbreak of the novel Coronavirus (COVID-19 or 2019-nCoV or SARS-CoV-2), in hindsight, one can state that the health authorities outside of China could have been better prepared in advance. With very little information coming out of an authoritarian China in the early days of the infection, social media could potentially have been used to look for early warning signs. Even after the spread of the disease, governments and communities, especially in the Western world, ignored the severity

✉ Amit Mittal
amit.mittal@chitkara.edu.in; yashamitmittal@gmail.com

Ruchi Mittal
ruchi.mittal@chitkara.edu.in

Ishan Aggarwal
ishan.aggarwal.1423@gmail.com

¹ Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India

² Chitkara Business School, Chitkara University, Rajpura, Punjab, India

³ Business Unit of Systems Support R&D, Ericsson India Global Services Pvt. Ltd., Gurugram, India

of the situation. The consequences were disastrous with the consequence that reported COVID-19 infections in the USA have overtaken infections in China and related deaths in Italy have surpassed China, as of March 27, 2020.

Epidemics and the spread of infectious diseases are known to have a significant negative socioeconomic impact due to strain on a society's resources and the rendering of the workforce less productive (Klepser 2014). The spread of infectious diseases is also known to spread xenophobia and social ostracism (Rich 2020). Recent outbreaks have shown very depressing images of people dying on the streets, empty tourist sites, empty airports and overwhelmed hospitals. For public health agencies, it is important not to work in isolation but they must work in close partnership with other stakeholders, especially the public and media (Chew and Eysenbach 2010). The news—real or fake—related to these outbreaks of highly infectious diseases coming out in media also plays a role in influencing public perception. This media can be broadly classified as formal or informal and mass media or social media. Most social media messages are informal and reflect the general mood and reactions of the public toward an important phenomenon (Al-Surimi et al. 2016). Starting December 2019, the world has been grappling with the new Coronavirus pandemic known as COVID-19. As of March 18, 2020, according to the Coronavirus Resource Centre data published online by John Hopkins University (<https://coronavirus.jhu.edu/>) there were 204,255 cases worldwide and 8243 deaths. The first known case was traced to Wuhan in China and the epicenter of this WHO declared pandemic has since shifted from China to Europe with countries such as Italy and Spain facing an unprecedented healthcare crisis. As of March 22, 2020, the total COVID-19-related deaths in Italy surpassed the total COVID-19-related deaths in China. One grim reality learned from the devastated northern part of Italy is that infectious diseases spread extremely fast once the existing healthcare system reaches its saturation point.

The application of data mining and big data analytics in the healthcare sector is prevalent, and several studies have earlier focused on epidemiological forecasting models for fatal and infectious diseases (Epstein et al. 2008; Hunter et al. 2017). Monitoring of social media, especially Twitter, has been very effective to capture the trend of disease outbreaks, specifically influenza (Culotta 2014). In earlier studies that have focused on forecasting through analytics on the impact of healthcare emergencies, modeling of the society has formed an important antecedent of the predictive model (Crooks and Hailegiorgis 2014; Aleman et al. 2011). Within the context of a society, and how it can potentially influence the spread of information related to the outbreak of a disease, sentiment analysis of messages posted on social media can provide critical insight into human perceptions and reactions toward the event. Furthermore, public panic

can potentially be reduced through the judicious use of media in the event of an environmental disaster (Yu et al. 2017) and the same may be inferred about epidemics. It is critical to stemming the flow of fake news and rumors to help control any infectious disease. Fake news also leads to social stigmatization and xenophobia as was recently seen toward Chinese people in various countries. The hashtag #ChineseDon'tComeToJapan was trending in Japan at one point in time Shimizu 2020). Gauging public opinion can help healthcare agencies and regulators design targeted communication campaigns which that can help stem the spread of rumors and unsubstantiated claims. This technique is very effectively used by political organizations during canvassing for elections with the 2012 Barack Obama presidential campaign's use of social media as a case in point (Ozturk and Ayvaz 2018).

One method to measure public sentiment on social media is sentiment analysis. Twitter is free and Twitter expressions are considered natural, spontaneous and heterogeneous (Gaspar et al. 2014) as a result of which public sentiment using Twitter data has been used in a variety of situations, e.g., infectious diseases (see Table 2); politics (Bose et al. 2019); Brexit (Agarwal et al. 2018); terrorism and radicalization (Macnair and Frank 2018; Iskander 2017); financial markets (Daniel et al. 2017; Pagolu et al. 2016); sports (Lucas et al. 2017); travel and tourism (Alaei et al. 2019; Park et al. 2016); refugee crisis (Ozturk and Ayvaz 2018); brand management (Liu et al. 2017); health care (Zhang et al. 2018); adoption of mobile apps (Pai and Alathur 2018); and disaster management (Panagiotopoulos et al. 2016) among a host of other areas. Sentiment analysis is based on the aggregation of the feelings or expressions of several individuals. In the context of a novel yet risky phenomenon, individuals are motivated to share information in a social system (such as social media) to (a) reduce personal perceived risk, (2) respond to others in the social system or (3) share new information with the social system thereby performing their "societal duty" (Gigerenzer and Selten 2002). The spread of new information by individuals or groups can also be seen in terms of "diffusion of innovation" (Rogers 2010). This diffusion of new information can also potentially lead to exaggeration or distortion within the social system. Controlling such new information can reduce the spread of fake news and misinformation (Bovet and Makse 2019).

In a study carried out to assess the measure of concern (MOC) among social media users concerning infectious diseases, Ji et al. (2015) conducted a sentiment analysis of tweets that were "personal" in nature. The present study seeks to conduct sentiment analysis based on polarity or affective valence of the sentiments generated on social media (Twitter) during the new Coronavirus (COVID-19) pandemic which was initially reported in Hubei province, Central China. Noteworthy, context is important in sentiment

analysis and is known to influence the results (Blitzer et al. 2007). The affective valence may be characterized as “positive,” “negative” and “neutral” (Gaspar et al. 2016). Based on the COVID-19 outcomes, this study shall seek to understand how internet users make sense of this unexpected event which has taken the shape of a major global healthcare crisis and which is still dynamic (see Table 1) and unfolding. The COVID-19 outcomes defined in the study are infections, deaths and recoveries. Specifically, this study seeks to resolve the following research question:

RQ What is the association between COVID-19 outcomes (infections, deaths and recoveries) and a tweet category based on affective valence (positive, negative and neutral)? Additionally, how do COVID-19 outcomes influence the distribution of tweet categories?

Table 1 Timeline of the COVID-19 outbreak. *Source:* <https://www.worldometers.info/coronavirus/>

Event	Date
First COVID-19 case reported in China	December 1, 2019
First COVID-19 case reported outside China	January 13, 2020 (in Thailand)
No new infection reported from Wuhan—ground zero	March 19, 2020
Epicenter shifts from China to Europe	March 10, 2020
WHO declares COVID-19 as a global pandemic	March 11, 2020
Top five countries with the highest number of reported COVID-19 infections:	
(1) China, (2) Italy, (3) USA, (4) Spain, (5) Germany	March 26, 2020
(1) USA, (2) Italy, (3) China, (4) Spain, (5) Germany	March 27, 2020

The subsequent sections include a literature review followed by the methodology applied. This is followed by data analysis and results, discussions and finally the conclusion.

2 Literature review

Health epidemics in general typically generate bursts of attention on social media, and the spread of infectious diseases has been a frequent topic of discourse specifically on Twitter (Table 2). Oyeyemi et al. (2014) discovered that during the Ebola virus epidemic in West Africa, a majority of Ebola-related tweets were spreading misinformation, for instance, drinking and washing in salty water as a cure. This not only complicated the management of the epidemic but also led to more deaths due to unverified claims and the spread of superstition. A separate study using Twitter data during the yellow fever outbreak in Angola, Congo and later in Kenya, China and Brazil in 2015 and 2016, also observed the spread of fake cures and misinformation through Twitter (Ortiz-Martínez and Jiménez-Arcia 2017). Ironically, despite the rampant poverty in that part of the world, the ownership of smartphones and access to the internet has rapidly increased. Lazard et al. (2015) in their study on the discovery of the first case of Ebola in the USA discuss how the CDC (Centers for Disease Control and Prevention) during a live chat sought to quell public panic and concerns through a live Twitter chat. Text mining of the chat data generated several themes that reflected the public’s concern, i.e., safe travel, virus lifespan, virus contraction and precautions. A study by (Ahmed et al. 2017) discovered that some social media users on Twitter shared humor and sarcasm related to Ebola. A thematic study of Twitter data related to the 2009 H1N1 (Swine flu) outbreak (Ahmed et al. 2019) offered some novel insights into how people communicate on social media. Their in-depth analysis led to the emergence of six broad

Table 2 Pandemics in the age of Twitter. *Source:* Authors compilation

Disease	Study
Ebola	Oyeyemi et al. (2014); Lazard et al. (2015); Ahmed et al. (2017); Guidry et al. (2017)
Swine Flu (H1N1)	Ahmed et al. (2019); Huo and Zhang (2016); Signorini et al. (2011); Chew and Eysenbach (2010); Ritterman et al. (2009)
Influenza-like illness (ILI)	Velardi et al. (2014); Achrekar et al. (2011)
Bird Flu (H7N9)	Hellsten et al. (2019); Broniatowski et al. (2013)
MERS	Fung et al. (2018); Kim et al. (2017); Shin et al. (2016)
SARS	St Louis and Zorlu (2012)
ZIKA	Pruss et al. (2019); Daughton and Paul (2019); Wirz et al. (2018)
COVID-19 (SARS-CoV-2)	Jahanbin and Rahmanian (2020); Chen et al. (2020a, b); Imran et al. (2020); Alamoodi et al. (2020)
Escherichia coli outbreak	Diaz-Aviles and Stewart (2012)
Measles	Tang et al. (2018); Radzikowski et al. (2016)
Yellow fever	Ortiz-Martínez and Jiménez-Arcia (2017)

themes related to the outbreak namely, (1) general fear, (2) naming H1N1 as “Swine” flu, (3) criticism of media, (4) racism and references to Mexicans, (5) consumption of pork and (5) popular culture and imminent apocalypse. In another earlier study on H1N1, Signorini et al. (2011) described how Twitter data, in addition to gauging public opinion, can also be used to estimate disease activity in real time. When the word spread that the symptoms are “mild,” buzz related to certain medications petered out on Twitter. A similar semantic approach to the analysis of Twitter sentiment was followed by Tang et al. (2018) during the 2015 outbreak of measles in the USA. Velardi et al. (2014) used a “symptom-driven” rather than “disease-driven” approach to analyzing Twitter data thereby offering a novel approach to predict the onset of infectious disease, specifically flu-like illnesses. Hellsten et al. (2019) used the media and communications issue arena framework to analyze bird flu discussions by various actors on Twitter. The context was the impact of bird flu on poultry farming in the Netherlands from 2015 to 2017. The study sought to differentiate between passive and active actors in the Twitter debate on the topic. (Fung et al. (2018) in a study covering the Middle East respiratory syndrome (MERS) which spread in South Korea in 2015, postulated that people with different linguistic and cultural backgrounds are likely to react differently to the disease on social media. Capturing MERS-related Twitter data in five languages, i.e., English, Korean, Japanese, Indonesian and Thai, they found their hypothesis is supported. People from different linguistic backgrounds did respond and react differently to the disease on social media. Pruss et al. (2019) used the conceptualization of differences in linguistic and language backgrounds to conduct a study in North America and South America where tweets in English, Spanish and Portuguese were extracted and analyzed in the context of the ZIKA virus infection. In another study on MERS, Kim et al. (2017) used Twitter data based on memes to identify the spatiotemporal patterns and risk assessment exhibited by social media users. St Louis and Zorlu (2012) examined how public health experts were exploiting social media resources to seek relevant information that may further be used to predict and manage highly infectious diseases such as SARS.

One very important theory that explains how individuals tend to share their perceived risk related to any phenomenon through informal and formal channels is the social amplification of risk framework (SARF; Kasperson et al. (1988). This information is further amplified—or weakened or distorted—and transmitted to other individuals or groups (Yoo et al. 2018; Chew and Eysenbach 2010). SARF has also been applied in the context of the spread of the ZIKA virus first reported in the spring of 2015 in Latin America (Wirz et al. 2018). In another recent study on the ZIKA virus outbreak, Daughton and Paul (2019) used Twitter data to identify the protective behavior exhibited by people especially

concerning their travel plans. The demographic and temporal correlates of such people were also measured. Twitter and other social media data were also used in a study that assessed how people coped in the scenario of a food crisis (Gaspar et al. 2014). The study was based on the *Escherichia coli* bacteria outbreak in Europe and the territory under study was Spain.

The spread of information about COVID-19 on Twitter presents an opportunity to understand public opinion in a scenario where the disease is still an ongoing phenomenon. Keeping this in mind, this study is unique as very few studies have been conducted using sentiment analysis of Twitter data during the early stages of COVID-19. This study shall be based on *sentiment analysis* (SA) which can be defined as a process that automates the mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through natural language processing (Danneman and Heimann 2014). SA can be considered as a tool to convert the raw opinions and ideas of millions of social media posts into valuable assets (Alamoodi et al. 2020). Sentiment analysis can cluster or classify social media posts based on polarity, i.e., polarity (negative, positive, neutral; see Gaspar et al. 2016). Advanced SA techniques can identify subjectivity (fact and opinion) (Onyenwe et al. 2020), a variety of emotions or sentiments such as happiness and sadness (Ali et al. 2017; Gao et al. 2016), concern, surprise, disgust or confusion (Ji et al. 2016). Some SA programs can also identify the object of an expression in addition to its sentiment (e.g., the color is nice but it is too loud: this is “negative”; Sloan and Quan-Haase 2017, p-546).

SA has also been used in a variety of contexts, with epidemics (see Table 2) being just one of them. Noteworthy, context is important in SA and is known to influence the results (Blitzer et al. 2007; Mittal et al. 2021). SA has been used in the context of customer feedback, elections, stock markets, sports, public protests, music concerts, terror incidents and disasters among a host of others. In this study, Twitter sentiment on COVID-19 using different relevant search terms or keywords shall be classified in terms of “positive” sentiments, “negative” sentiments and “neutral” sentiments, which together are termed as affective valence.

3 Method

There are various methodologies to identify textual sentiments. These methodologies are mainly classified into Lexicon-based and machine learning (Alamoodi et al. 2020). Lexicon-based methods focus on deriving opinion on the input text and determine positive or negative or neutral sentiments from the input text and assigning it a polarity value. Lexicon-based methods make use of predefined dictionaries of words, in which each word is

associated with a specific sentiment. In machine learning techniques, a sentiment analysis model is built based on a training process populated with data labeled by humans (Messias et al. 2017). This study uses the Lexicon-based approach, which is considered very effective for sentiment analysis (Chen et al. 2020a, b). After exploring the various tools, it was decided to go ahead with nltk.sentiment.vader sentiment intensity analyzer module available from Python. VADER has been extremely successful in dealing with social media content and editorial reviews. This is because of the algorithms used by VADER which not only categorize the input text into positive or negative or neutral categories, but all compute the compound polarity scores to gain a fair idea about how positive or negative the sentiments are. VADER is applied not only for social network data but also on other data generated online such as emails (Borg and Boldt 2020) and this tool remains very popular, effective and contemporary (Dahal et al. 2019; Moutidis and Williams 2020; Wei et al. 2016). VADER has a lot of advantages -

- It works very well with social media content, movies, products and editorial reviews.
- There is no need for any training data. It gets formed on the basis of valence, standard sentiment lexicons.
- It can even be used with streaming data online.
- It is quite fast and there is no need to compromise on the speed performance.

This study is largely based on descriptive algorithms and the process consisted of four steps (figure 1):

- Data Acquisition** Fetched Twitter Remote Data using web service *tweepy*. It is a standard python library for accessing the Twitter API and used search () API call to match the specified query and used wait_on_rate_limit flag (with the value set to True) to automatically wait for rate limits to replenish. *jsonpickle*: It is again a python library used for serialization and deserialization purposes. It transforms complex Python objects to and from JSON and used encode() API call to transform the object into *json* string. The keywords used to extract the data were: “Coronavirus,” “COVID-19,” “Wuhan,” “Pandemic” and “Corona.” Only tweets in English were captured.
- Data Preparation** This has two components. Firstly, exploration was conducted using *json* which performed data modeling and transformed the data into a semi-structured format. This was followed by data preprocessing: (a) removed hashtags, mentions, (b) removed invalid data with indefinite spaces and (c) filtered duplicate records based on tweet ID.
- Data Analysis** As mentioned earlier, for classification, NLTK and VADER were used. Natural Language Tool Kit Vader python module helps in sentiment analysis of data in human language. VADER is used for sentiment analysis on the basis of lexicon and rule-based approach. VADER library makes use of various combinations (mostly on the basis of handling emojis, slangs, emoticons, degree modifiers, capitalizations, punctuations and usage of conjunctions) for computing the polarity score of sentences to categorize them as positive or negative or neutral. Automated Coding was an added functionality to identify re-tweets and

Data Analysis Model

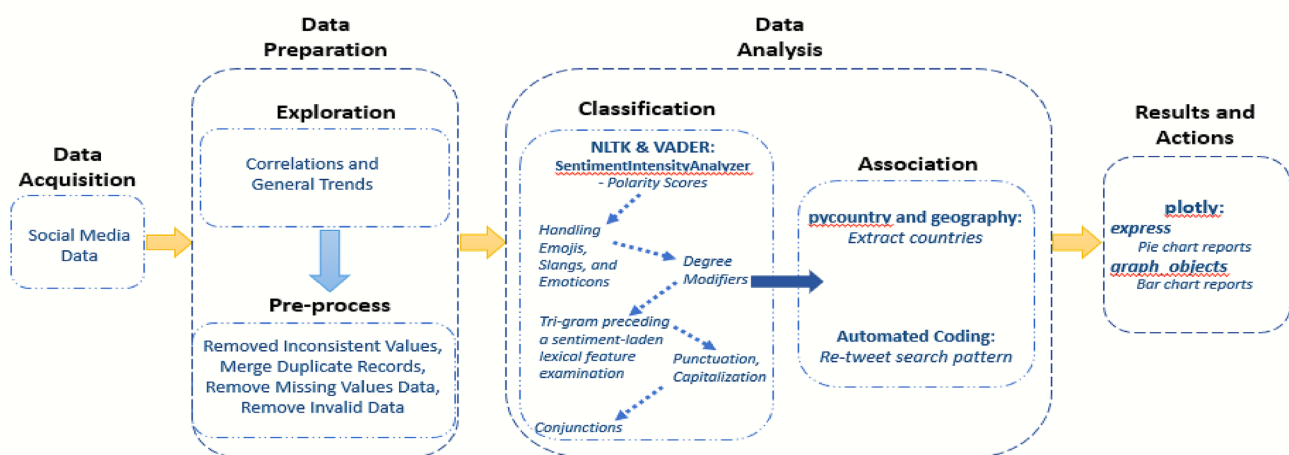


Fig. 1 Data analysis framework

the number of re-tweets for positive, negative and neutral sentiments out of the total tweets.

- iv. *Results and Actions* Plotly was used where Plotly Express generated a terse, consistent, high-level API for rapid data exploration and figure generation. Similarly, Plotly graph_objects generated the graph object figures with any of the named attributes.

An example of a processed tweet is given as per exhibit 1.

Approximately 50,000 tweets per day were extracted between the period March 1, 2020–March 24, 2020. Besides, 70,095 tweets were extracted for a week in March 2020 to run a pilot data analysis (see Table 3).

4 Data analysis and results

Using the coding-based sentiment analysis, three categories of tweets and re-tweets were extracted based on affective valence (see table 3) namely positive tweets, negative tweets and neutral tweets. This was a preliminary extraction—a pilot study –on which processing was done to confirm whether a sentiment analysis can be done on the data in line with the objectives. This was similar to the approach followed by Agarwal et al. (2011). Neutral tweets seem

Table 3 Random tweets extracted from March 16, 2020–March 23, 2020

Category of tweets	Total tweets (Percentage of Total)
Total tweets captured	70,095
Positive tweets	23,057 (33%)
Negative tweets	17,471 (30%)
Neutral tweets	21,945 (37%)

to be more in numbers compared to positive or negative tweets reflecting the general indifference or “non-alarmist” approach of this category of social media users in the duration when the tweets were extracted. The examples of how “positive” tweets, “negative” tweets and “neutral” tweets in the context of COVID-19 are described as given in Table 4.

Exhibit 1: Stages of processing a raw tweet (example)

Original tweet:
 We stand by Italy during these trying times. Share your Support for our Italian friends, They are our colleagues, friends, and family. 🇮🇹 🇮🇹 🇮🇹 🇮🇹 Cari amici, siamo con voi.#COVID-19 #WeStandWithItaly

Tweet content extracted and saved:
 We stand by Italy during these trying times. Share your Support for our Italian friends, They are our colleagues, friends, and family.#COVID-19 #WeStandWithItaly

Preprocessing of original tweet:
Remove links:
 We stand by Italy during these trying times. Share your Support for our Italian friends, They are our colleagues, friends, and family. #COVID-19 #WeStandWithItaly

Remove hashtags, mentions and emojis:
 We stand by Italy during these trying times. Share your Support for our Italian friends, They are our colleagues, friends, and family

Spell checker:
 We stand by Italy during these trying times share your support for our Italian friends they are our colleagues friends and family

Cleaned tweet:
 “We stand by Italy during these trying times share your support for our Italian friends they are our colleagues, friends, and family...”

As given in table 5, the figures for COVID-19 reported cases of infection and deaths are figures for the world, while the figures for COVID-19 reported cases of recoveries are for China. Prior research identifies perceptions as a precursor to sentiments (Myslín et al. 2013; Boon-Itt and Skunkan

Table 4 Description and samples of tweet categories

Positive tweets	“We are in this together though apart. Helping people remain at home will save lives and keep people safe...” “USA stands by Italy during these trying times Share your Support for our Italian friends, they are our colleagues and friends...” “Best wishes for a speedy recovery and take care South Korea and have a great success through testing...”
Negative tweets	“The will be the scapegoat The US economy was headed for a crash anyway due to the reckless and incompetent White House...” “It is expected at least the same amount of people to die to this vile rancid virus....The least we can do whilst we still can...” “IRONY IS DEAD Mike Pompeo just criticized China for hiding the Coronavirus from us”
Neutral tweets	“US virologist on COVID- This will end but expect more” “XYD Clinic Labs ramp up for round-the-clock COVID- testing” “Well.... there goes my wedding..”

Table 5 Day-wise (23 days in March 2020), negative and positive tweets and the number of COVID-19 infected people across the world, related deaths and recoveries (in China). *Source:* for (a) and (b): Extracted by the authors; for (c) and (d) https://covid.ourworldindata.org/data/ecdc/total_cases.csv accessed on 24.03.2020 and for (e)

Date	Positive tweets (a)	Negative tweets (b)	Neutral tweets (c)	Infections (c)	Deaths (d)	Recoveries (e)
1-Mar-20	18,420	13,966	17,214	87,024	3050	44,462
2-Mar-20	12,020	11,917	25,116	89,068	3117	47,204
3-Mar-20	10,520	12,354	11,234	90,663	3202	49,856
4-Mar-20	10,113	10,250	29,540	93,076	3285	52,045
5-Mar-20	15,889	12,281	14,630	95,315	3387	53,726
6-Mar-20	12,030	14,281	13,687	98,171	3494	55,404
7-Mar-20	16,346	15,366	14,286	102,132	3599	57,065
8-Mar-20	10,500	14,500	13,004	105,823	3827	58,600
9-Mar-20	10,132	13,981	25,843	109,694	4025	59,897
10-Mar-20	20,080	16,010	13,820	114,231	4296	61,475
11-Mar-20	10,850	14,516	12,544	118,609	4628	62,793
12-Mar-20	15,892	19,250	13,658	125,496	4981	64,111
13-Mar-20	11,003	15,421	14,110	133,848	5429	65,541
14-Mar-20	20,112	12,118	17,010	143,223	5833	66,911
15-Mar-20	16,524	14,992	15,444	151,363	6519	67,749
16-Mar-20	12,902	15,258	21,745	167,414	7161	68,679
17-Mar-20	17,032	15,772	17,123	180,159	7978	69,601
18-Mar-20	16,838	15,658	17,010	194,909	8951	70,420
19-Mar-20	29,069	15,204	3445	213,254	10,031	71,150
20-Mar-20	27,445	16,753	6407	242,473	11,387	71,740
21-Mar-20	16,490	14,594	17,900	271,228	13,013	72,244
22-Mar-20	17,620	16,200	15,010	305,275	14,641	72,703
23-Mar-20	12,755	18,569	17,009	338,303	16,514	73,159

The graphical representation are described as per Figs. 2 and 3

2020; Blasi et al 2020). The reason for using figures from China is based on the premise that Twitter users’ sentiments would have been influenced by their perceptions toward the news of COVID-19 recoveries (regardless of their accuracy) in China which was the initial epicenter. Furthermore, in the early days, only China experienced significant recoveries after the infection due to the recency of the disease. Hence, “COVID-19 recoveries” in China was used as a surrogate variable. The day-wise distribution of tweets and COVID-19 infections are given as given in Table 5. Based on the data extracted the following tests were performed:

1. Measure the association between total infections, deaths and recoveries and positive, negative and neutral tweets. *Test performed: correlation matrix based on Pearson’s correlation.*
2. Measure significant differences between COVID outcomes (infections, deaths and recoveries) and tweets, i.e., positive, negative and neutral. *Test performed: one-way ANOVA*

Pearson correlation was applied to the data to measure the association between different variables. Pearson correlation

is calculated as follows where r is the correlation coefficient and x and y are the variables under study.

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

Table 6 shows that there is a significant association between total global cumulative infections and negative tweets, between global cumulative deaths and negative tweets, and even between cumulative recoveries (in China) and negative tweets. On the other hand, this was no significant association between total global cumulative infections and positive tweets and between global cumulative deaths and positive tweets. However, there was a significant positive association between cumulative recoveries (in China) and positive tweets. Additionally, Table 7 shows that the positive tweets outnumber the negative tweets and neutral tweets outnumber the positive tweets. In the context of the period captured, this shows the inherent optimism or indifference in people despite such a depressing scenario.

One-way ANOVA has been applied to measure whether there is any significant difference between COVID outcomes

Fig. 2 COVID-19 outcomes

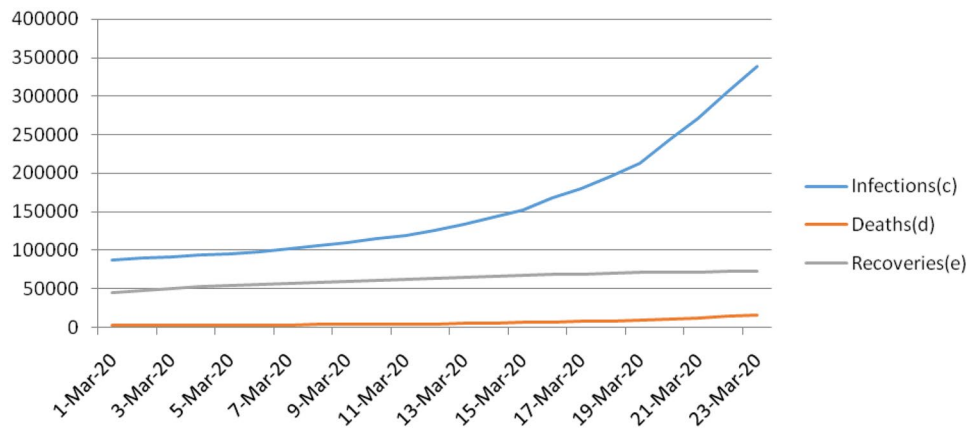


Fig. 3 Tweet categories

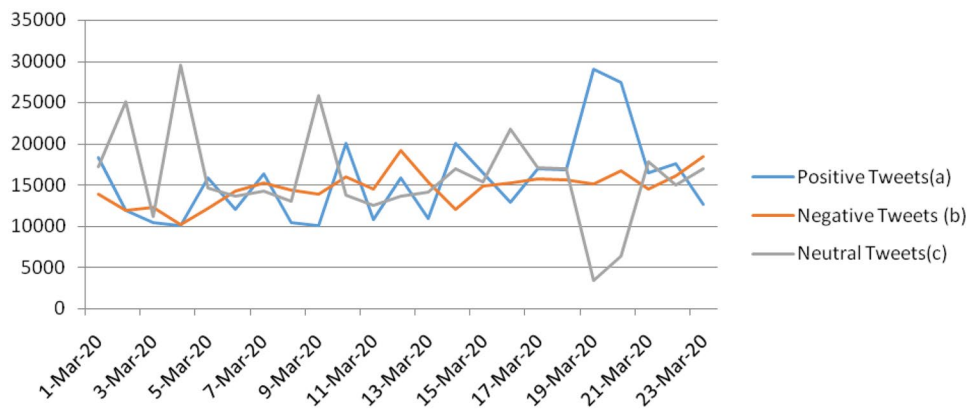


Table 6 Correlation matrix: association between tweet category and COVID-19-related cases

	Infections	Deaths	Recoveries	Positive tweets	Negative tweets	Neutral tweets
Infections	1					
Deaths	0.996**	1				
Recoveries	0.817**	0.814**	1			
Positive tweets	0.379	0.388	0.417*	1		
Negative tweets	0.559**	0.555**	0.628**	0.292	1	
Neutral tweets	-0.084	-0.086	-0.213	-0.389	-0.435*	1

*Correlation is significant at the 0.05 level (two-tailed) **correlation is significant at the 0.01 level (two-tailed)—technique used is Pearson correlation coefficient

Table 7 Descriptive statistics

Variables	N	Minimum	Maximum	Mean	Std. deviation
Positive count	23	10,113	29,069	15,677	5118.81
Negative count	23	10,250	19,250	14,748	2071.08
Neutral count	23	3445	29,540	15,947	5737.51

and the distribution of tweets in three different categories namely: positive, negative and neutral. The results (Table 8) highlights that there are significant differences in all the COVID outcomes across different tweet categories.

However, it is also important to note that infections and deaths are significant at a 0.05% level. On the other side, recovery is significant at a 0.10% level which is comparatively lower than infections and death.

5 Discussions

Public health surveillance is critical to deploying resources in the event of the spread of a highly infectious disease (Ji et al. 2015). Public health communication directed to concerned stakeholders could be strengthened through the

Table 8 ANOVA for COVID outcomes and tweet counts

COVID outcomes	Sum of squares	Df	Mean square	F	Sig.
<i>Infections</i>					
Between groups	657,357,828.193	2	328,678,914.096	0.056	0.046**
Within groups	117,913,285,052.764	20	5,895,664,252.638		
Total	118,570,642,880.957	22			
<i>Deaths</i>					
Between groups	2,052,770.749	2	1,026,385.374	0.059	0.034**
Withi groups	348,784,384.556	20	17,439,219.228		
Total	350,837,155.304	22			
<i>Recoveries</i>					
Between groups	14,791,553.359	2	7,395,776.680	0.088	0.916***
Within groups	1,676,724,071.597	20	83,836,203.580		
Total	1,691,515,624.957	22			

***Significant at 1.0% level, **Significant at 0.05% level

effective leveraging of social media. The explosive growth in the number of smartphones and increasing reach of the internet globally has led to several individuals altogether abandoning or reducing their usage of traditional media in favor of online sources, and this provides public health institutions with an opportunity to use social media to their advantage (Olusola et al. 2017; Smith and Brenner 2012). To leverage the communicative power of social media, it is important to understand the sentiments of people (Twitterati) in the virtual world. As per the results generated through sentiment analysis of tweets in this study, the emergence of “positive” tweets (Table 3) in such a gloomy pandemic scenario shows the inherent resilience of humans. The significant association between positive tweets and news of COVID-19 recoveries seems to hint at a more optimistic scenario whenever the pandemic finally comes to an end or is controlled. Such public reactions—for good—have the potential to go viral and influence several others, especially those who are classified as “neutral” or fence-sitters.

The findings are important for healthcare agencies as people need to remain optimistic and not panic. In such pandemic situations where millions of people worldwide are in a lockdown situation, people need to maintain their mental health as well and not panic (Alamoodi et al. 2020). As a case in point, researchers in the University of Pennsylvania (USA) created a dashboard to track and analyze changes in language expressed on Twitter during the COVID-19 pandemic in the USA, with a focus on mental health and symptom mentions (Guntuku et al. 2020).

In a study published in the American Journal of Infection Control, it was discovered that even before the official announcement of the outbreak of Ebola in West Africa, news about its spread had already appeared in tweets three days in advance (Odlum and Yoon. 2015). These tweets reached 60 million social media users in these three days. In the context of COVID-19-related social media chatter in the

USA, topic detection and sentiments on Twitter identified several disturbing trends as well, e.g., vaccine opposition and COVID-19 as a conspiracy (Jamison et al. 2020), cyber racism (Dubey 2020), distrust of health authorities (Bonnievie et al. 2021).

This shows the communicative power (negative as well as positive) of social media especially Twitter. The study by McCombs and Shaw (1972)—which discusses how media can influence the salience of issues—seen in conjunction with the social amplification theory (Kasperson et al. 1988)—which discusses how people spread risk perception through media—shows that social media platforms such as Twitter present a formidable challenge and a great opportunity as well. In fact, in the era of fake news (Albright 2017), scanning social media becomes indispensable.

6 Conclusions

This study is an attempt to conduct text mining of affective valence of the sentiments generated on Twitter during the new Coronavirus (COVID-19) pandemic. The study utilized Twitter data generated during March 2020 due to the inherent advantages of using Twitter data, i.e., the data are real time, free, brief and have immense reach irrespective of national boundaries. In March 2020, the pandemic was still very new and opinions were still shaping. Government and organizational policies as to how to deal with the disease were still tentative and evolving and public opinion likewise was divided. This study extracted tweets that were characterized as “positive,” “negative” and “neutral” to assess how this categorization may relate to COVID-19 outcomes, i.e., the number of infections, deaths and recoveries. Understanding public sentiments may help shape policies and procedures for the government and other organizations given that

the true impact of the disease shall be felt only when the pandemic ends (Hamzah et al. 2020).

In terms of future research, this study could be extended beyond the affective valence of tweets in terms of positive, negative or neutral and conduct a qualitative analysis of the tweets in terms of themes or other qualitative parameters. There are many more emotions that can be captured. Even during the pandemic, different aspects need to be studied. For instance, the social media response to how the pandemic is impacting specific groups such as migrants, healthcare workers, people deputed on essential duties (for example, Police and Public Administrators). Additionally, tweets in other languages could be captured and analyzed especially given that several non-English speaking countries are currently in the grip of COVID-19.

Authors' contributions AM was responsible for the manuscript proof-reading, language editing, part of the data analysis and identification of past literature. IA extracted the data and performed the sentiment analysis. RM wrote major parts of the manuscript.

Declarations

Conflict of interest We declare no conflict of interests whatsoever.

References

- Achrekar H, Gandhe A, Lazarus R, Yu SH, & Liu B (2011). Predicting flu trends using Twitter data. In: 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPs). IEEE. pp. 702–707
- Agarwal A, Singh R, Toshniwal D (2018) Geospatial sentiment analysis using twitter data for UK-EU referendum. *J Inf Optim Sci* 39(1):303–317
- Agarwal A, Xie B, Vovsha I, Rambow O, & Passonneau RJ (2011). Sentiment analysis of twitter data. In Proceedings of the workshop on language in social media (LSM 2011). pp. 30–38
- Ahmed W, Bath PA, Sbaffi L, Demartini G (2019) Novel insights into views towards H1N1 during the 2009 pandemic: a thematic analysis of Twitter data. *Health Info Libr J* 36(1):60–72
- Ahmed W, Demarini G, & Bath PA (2017) Topics discussed on twitter at the beginning of the 2014 Ebola epidemic in United States. In: Conference 2017 proceedings
- Alaei AR, Becken S, Stantic B (2019) Sentiment analysis in tourism: capitalizing on big data. *J Travel Res* 58(2):175–191
- Alamoodi A, Zaidan B, Zaidan A, Albahri O, Mohammed K, Malik R, Hameed H (2020) Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review. *Expert Syst Appl*. <https://doi.org/10.1016/j.eswa.2020.114155>
- Albright J (2017) Welcome to the era of fake news. *Media Commun* 5(2):87–89
- Aleman DM, Wibisono TG, Schwartz B (2011) A nonhomogeneous agent-based simulation approach to modeling the spread of disease in a pandemic outbreak. *Interfaces* 41(3):301–315
- Alowaidi S, Saleh M, Abulnaja O (2017) Semantic sentiment analysis of Arabic texts. *Int J Adv Comput Sci Appl* 8(2):256–262
- Al-Surimi K, Khalifa M, Bahkali S, Ashraf EM, & Househ M (2016). The potential of social media and internet-based data in preventing and fighting infectious diseases: from internet to Twitter. In: Emerging and re-emerging viral infections. Springer, Cham. pp. 131–139
- Ali K, Dong H, Bouguettaya A, Erradi A, & Hadjidj R (2017). Sentiment analysis as a service: a social media based sentiment analysis framework. In 2017 IEEE international conference on web services (ICWS), pp. 660–667
- Blasi S, Brigato L, Sedita SR (2020) Eco-friendliness and fashion perceptual attributes of fashion brands: an analysis of consumers' perceptions based on twitter data mining. *J Clean Prod* 244:118701
- Blitzer J, Dredze M, & Pereira F (2007) Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In: Proceedings of the 45th annual meeting of the association of computational linguistics. pp. 440–447
- Bonnevie E, Gallegos-Jeffrey A, Goldberg J, Byrd B, Smyser J (2021) Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *J Commun Health* 14(1):12–19
- Boon-Itt S, Skunkan Y (2020) Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health Surveill* 6(4):e21978
- Borg A, Boldt M (2020) Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Syst Appl* 162:113746
- Bose R, Dey RK, Roy S, & Sarddar D (2019) Analyzing political sentiment using Twitter data. In Information and communication technology for intelligent systems. Springer, Singapore. pp. 427–436
- Bovet A, Makse HA (2019) Influence of fake news in Twitter during the 2016 US presidential election. *Nat Commun* 10(1):1–14
- Broniatowski DA, Paul MJ, Dredze M (2013) National and local influenza surveillance through Twitter: an analysis of the 2012–2013 influenza epidemic. *PLoS one* 8(12):e83672
- Chen E, Lerman K, & Ferrara E (2020) COVID-19: the first public coronavirus twitter dataset. arXiv preprint <http://arxiv.org/abs/2003.07372>
- Chen J, Hossain MS, Zhang H (2020b) Analyzing the sentiment correlation between regular tweets and retweets. *Soc Netw Anal Min* 10(1):13
- Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS one* 5(11):e14118
- Crooks AT, Hailegiorgis AB (2014) An agent-based modeling approach applied to the spread of cholera. *Environ Model Softw* 62:164–177
- Culotta A (2014) Estimating county health statistics with twitter. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp 1335–1344
- Dahal B, Kumar SA, Li Z (2019) Topic modeling and sentiment analysis of global climate change tweets. *Soc Netw Anal Min* 9(1):24
- Daniel M, Neves RF, Horta N (2017) Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Syst Appl* 71:111–124
- Danneman N, & Heimann R (2014) Social media mining with R. Packt Publishing Ltd
- Daughton AR, Paul MJ (2019) Identifying protective health behaviors on Twitter: observational study of travel advisories and Zika virus. *J Med Internet Res* 21(5):e13090
- Diaz-Aviles E, & Stewart A (2012) Tracking twitter for epidemic intelligence: case study: Ehec/hus outbreak in germany, 2011. In Proceedings of the 4th annual ACM web science conference. pp. 82–85
- Dubey AD (2020) The resurgence of cyber racism during the COVID-19 pandemic and its aftereffects: analysis of sentiments and emotions in tweets. *JMIR Public Health Surveill* 6(4):e19833

- Epstein JM, Parker J, Cummings D, Hammond RA (2008) Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PLoS One* 3(12):e3955
- Fung ICH, Zeng J, Chan CH, Liang H, Yin J, Liu Z, Fu KW (2018) Twitter and middle east respiratory syndrome, South Korea, 2015: a multi-lingual study. *Infect, Dis Health* 23(1):10–16
- Gaspar R, Gorjao S, Seibt B, Lima L, Barnett J, Moss A et al (2014) Tweeting during food crises: a psychosocial analysis of threat coping expressions in Spain during the 2011 European EHEC outbreak. *Int J Hum Comput Stud* 72:239–254
- Gaspar R, Pedro C, Panagiotopoulos P, Seibt B (2016) Beyond positive or negative: qualitative sentiment analysis of social media reactions to unexpected stressful events. *Comput Hum Behav* 56:179–191
- Gao B, Berendt B, Vanschoren J (2016) Toward understanding online sentiment expression: an interdisciplinary approach with subgroup comparison and visualization. *Soc Netw Anal Min* 6(1):68
- Gigerenzer G, & Selten R (Eds.) (2002). *Bounded rationality: The adaptive toolbox*. MIT press
- Guidry JP, Jin Y, Orr CA, Messner M, Meganck S (2017) Ebola on Instagram and Twitter: how health organizations address the health crisis in their social media engagement. *Public Relations Review* 43(3):477–486
- Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, Ungar LH (2020) Tracking mental health and symptom mentions on twitter during covid-19. *J Gen Intern Med* 35(9):2798–2800
- Hamzah FB, Lau C, Nazri H, Ligot DV, Lee G, Tan CL (2020) CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Org* 1(32):264
- Hellsten I, Jacobs S, Wonneberger A (2019) Active and passive stakeholders in issue arenas: a communication network approach to the bird flu debate on Twitter. *Pub Rel Rev* 45(1):35–48
- Hunter E, Mac NB, Kelleher JD (2017) A taxonomy for agent-based models in human infectious disease epidemiology. *J Artif Soc Soc Sim*. <https://doi.org/10.18564/jasss.3414>
- Huo HF, Zhang XM (2016) Modeling the influence of Twitter in reducing and increasing the spread of influenza epidemics. *Springerplus* 5(1):1–20
- Imran AS, Daudpota SM, Kastrati Z, Batra R (2020) Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *IEEE Access* 8:181074–181090
- Jahanbin K, Rahmanian V (2020) Using twitter and web news mining to predict COVID-19 outbreak. *Asia Pacific J Trop Med* 13(8):378
- Jamison AM, Broniatowski DA, Dredze M, Sangraula A, Smith MC, Quinn SC (2020) Not just conspiracy theories: vaccine opponents and proponents add to the COVID-19 ‘infodemic’ on Twitter. *Harv Kennedy Sch Misinf Rev*. <https://doi.org/10.37016/mr-2020-38>
- Ji X, Chun SA, Wei Z, Geller J (2015) Twitter sentiment classification for measuring public health concerns. *Soc Netw Anal Min* 5(1):13
- Ji X, Chun SA, & Geller J (2016) Knowledge-based tweet classification for disease sentiment monitoring. In *Sentiment analysis and ontology engineering*. pp. 425–454
- Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R, Ratick S (1988) The social amplification of risk: a conceptual framework. *Risk Anal* 8(2):177–187
- Kim IH, Feng CC, Wang YC, Spitzberg BH, Tsou MH (2017) Exploratory spatiotemporal analysis in risk communication during the MERS outbreak in South Korea. *Prof Geogr* 69(4):629–643
- Klepser ME (2014) Socioeconomic impact of seasonal (Epidemic) influenza and the role of over-the-counter medicines. *Drugs* 74:1467–1479. <https://doi.org/10.1007/s40265-014-0245-1>
- Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M (2015) 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* 43(10):1109–1111
- Liu X, Burns AC, Hou Y (2017) An investigation of brand-related user-generated content on Twitter. *J Advert* 46(2):236–247
- Lucas GM, Gratch J, Malandrakis N, Szablowski E, Fessler E, Nichols J (2017) GOAALLL!: using sentiment in the world cup to explore theories of emotion. *Image Vis Comput* 65:58–65
- Macnair L, Frank R (2018) The mediums and the messages: exploring the language of Islamic State media through sentiment analysis. *Crit Stud Terrorism* 11(3):438–457
- McCombs M, Shaw D (1972) The agenda setting function of mass media. *Public Opin Q* 36(2):176–187
- Messias J, Diniz JP, Soares E, Ferreira M, Araújo M, Bastos L, Benvenuto F (2017) An evaluation of sentiment analysis for mobile devices. *Soc Netw Anal Min* 7(1):20
- Mittal R, Ahmed W, Mittal A, & Aggarwal I (2021) Twitter users exhibited coping behaviours during the COVID-19 lockdown: an analysis of tweets using mixed methods. *Inf Discov Deliv, Ahead-of-Print*
- Moutidis I, Williams HT (2020) Good and bad events: combining network-based event detection with sentiment analysis. *Soc Netw Anal Min* 10(1):1–12
- Myslín M, Zhu SH, Chapman W, Conway M (2013) Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res* 15(8):e174
- Odlum M, Yoon S (2015) What can we learn about the Ebola outbreak from tweets? *Am J Infect Control* 43(6):563–571
- Olusola A, Ibrahim S, Priscilla G (2017) An era of journalism transition in South Africa: traditional media versus online media. *J Soc Sci* 51(1–3):1–5
- Ombabi AH, Ouarda W, Alimi AM (2020) Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. *Soc Netw Anal Min* 10(1):1–13
- Ortiz-Martínez Y, Jiménez-Arcia LF (2017) Yellow fever outbreaks and Twitter: rumors and misinformation. *Am J Infect Control* 45(7):816–817
- Onyenwe I, Nwagbo S, Mbeledogu N, Onyedinma E (2020) The impact of political party/candidate on the election results from a sentiment analysis perspective using# AnambraDecides2017 tweets. *Soc Netw Anal Min* 10(1):1–17
- Oyeyemi SO, Gabarron E, Wynn R (2014) Ebola, Twitter, and misinformation: a dangerous combination? *Bmj* 349:g6178
- Öztürk N, Ayvaz S (2018) Sentiment analysis on Twitter: a text mining approach to the Syrian refugee crisis. *Telematics Inform* 35(1):136–147
- Pagolu VS, Reddy KN, Panda G, & Majhi B (2016) Sentiment analysis of Twitter data for predicting stock market movements. In: 2016 international conference on signal processing, communication, power and embedded system (SCOPEs). IEEE. pp. 1345–1350
- Pai RR, Alathur S (2018) Assessing mobile health applications with twitter analytics. *Int J Med Inform* 113:72–84
- Panagiotopoulos P, Barnett J, Bigdeli AZ, Sams S (2016) Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technol Forecast Soc Chang* 111:86–96
- Park SB, Ok CM, Chae BK (2016) Using Twitter data for cruise tourism marketing and research. *J Travel Tour Mark* 33(6):885–898
- Pruss D, Fujinuma Y, Daughton AR, Paul MJ, Arnot B, Szafir DA, Boyd-Graber J (2019) Zika discourse in the Americas: a multilingual topic analysis of Twitter. *PLoS one* 14(5):e0216922
- Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL (2016) The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health Surveill* 2(1):e1
- Rich M (2020) As coronavirus spreads, so does anti-Chinese sentiment. *The New York Times*, 30

- Ritterman J, Osborne M, & Klein E (2009) Using prediction markets and Twitter to predict a swine flu pandemic. In: 1st international workshop on mining social media 9. pp. 9–17
- Rogers EM (2010) Diffusion of innovations. Simon and Schuster
- Shimizu K (2020) 2019-nCoV, fake news, and racism. *The Lancet* 395(10225):685–686
- Shin SY, Seo DW, An J, Kwak H, Kim SH, Gwack J, Jo MW (2016) High correlation of middle east respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci Rep* 6:32920
- Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS one* 6(5):e19467
- Sloan L, & Quan-Haase A (Eds.) (2017) *The SAGE handbook of social media research methods*. Sage
- Smith A. & Brenner J (2012). Twitter use 2012. Pew internet & American life project, 4
- St Louis C, Zorlu G (2012) Can Twitter predict disease outbreaks? *Bmj* 344:e2353
- Szomszor M, Kostkova P, & St Louis C (2011) Twitter informatics: tracking and understanding public reaction during the 2009 swine flu pandemic. In: 2011 IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology 1. IEEE. pp. 320–323
- Tang L, Bie B, Zhi D (2018) Tweeting about measles during stages of an outbreak: a semantic network approach to the framing of an emerging infectious disease. *Am J Infect Control* 46(12):1375–1380
- Velardi P, Stilo G, Tozzi AE, Gesualdo F (2014) Twitter mining for fine-grained syndromic surveillance. *Artif Intell Med* 61(3):153–163
- Wei W, Joseph K, Liu H, Carley KM (2016) Exploring characteristics of suspended users and network stability on Twitter. *Soc Netw Anal Min* 6(1):51
- Wirz CD, Xenos MA, Brossard D, Scheufele D, Chung JH, Massarani L (2018) Rethinking social amplification of risk: social media and Zika in three languages. *Risk Anal* 38(12):2599–2624
- Yoo S, Song J, Jeong O (2018) Social media contents based sentiment analysis and prediction system. *Expert Syst Appl* 105:102–111
- Yu L, Li L, Tang L (2017) What can mass media do to control public panic in accidents of hazardous chemical leakage into rivers? A multi-agent-based online opinion dissemination model. *J Clean Prod* 143:1203–1214
- Zhang L, Hall M, Bastola D (2018) Utilizing Twitter data for analysis of chemotherapy. *Int J Med Inform* 120:92–100

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.