An "Infodemic": Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak

Richard J. Medford,[1,3] Sameh N. Saleh,[2,3,a] Andrew Sumarsono,[2] Trish M. Perl,[1] and Christoph U. Lehmann[3]

[1]University of Texas Southwestern Medical Center, Department of Internal Medicine, Division of Infectious Diseases and Geographic Medicine, Dallas, Texas, USA, [2]University of Texas Southwestern Medical Center, Department of Internal Medicine, Dallas, Texas, USA, [3]University of Texas Southwestern Medical Center, Clinical Informatics Center, Dallas, Texas, USA

Background. Twitter has been used to track trends and disseminate health information during viral epidemics. On January 21, 2020, the Centers for Disease Control and Prevention activated its Emergency Operations Center and the World Health Organization released its first situation report about coronavirus disease 2019 (COVID-19), sparking significant media attention. How Twitter content and sentiment evolved in the early stages of the COVID-19 pandemic has not been described.

Methods. We extracted tweets matching hashtags related to COVID-19 from January 14 to 28, 2020 using Twitter's application programming interface. We measured themes and frequency of keywords related to infection prevention practices. We performed a sentiment analysis to identify the sentiment polarity and predominant emotions in tweets and conducted topic modeling to identify and explore discussion topics over time. We compared sentiment, emotion, and topics among the most popular tweets, defined by the number of retweets.

Results. We evaluated 126 049 tweets from 53 196 unique users. The hourly number of COVID-19-related tweets starkly increased from January 21, 2020 onward. Approximately half (49.5%) of all tweets expressed fear and approximately 30% expressed surprise. In the full cohort, the economic and political impact of COVID-19 was the most commonly discussed topic. When focusing on the most retweeted tweets, the incidence of fear decreased and topics focused on quarantine efforts, the outbreak and its transmission, as well as prevention.

Conclusions. Twitter is a rich medium that can be leveraged to understand public sentiment in real-time and potentially target individualized public health messages based on user interest and emotion.

Keywords. COVID-19; pandemic; SARS-CoV-2; sentiment analysis; topic modeling.

With over 300 million monthly users, the microblogging platform Twitter is increasingly used to disseminate public health information and obtain real-time health data using crowdsourcing methods [1]. Researchers analyzed Twitter data to project the spread of influenza and other infectious outbreaks in real time [2]. In 2009, investigators measured the evolving interest in an influenza A outbreak by analyzing tweet keywords and estimating real-time disease activity and disease prevention efforts [3]. During the Ebola virus (EV) outbreak in 2014, Twitter users publicized pertinent health information from media sources with peak Twitter activity within 24 hours after news events [4]. Tweet content analysis after the EV epidemic discovered that Ebola-related tweets revolved mainly around risk factors, prevention, disease trends, and compassion [5]. Likewise, during the 2015 Middle Eastern respiratory syndrome outbreak, disease spread was found to be correlated with Twitter activity, promoting Twitter as a potential surveillance tool for emerging infectious diseases [6]. During the Zika virus epidemic, Twitter was used to study significant changes in travel behavior due to mounting public concerns [7]. Recognizing Twitter's potential to inform and educate the public, governmental agencies such as the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) have adopted the use of Twitter and other social media. In the first 12 weeks of the Zika outbreak in late 2015, the WHO Twitter account was retweeted over 20 000 times, demonstrating its widespread impact on disseminating health information [8].

In December 2019, the first diagnosis of a novel, emerging coronavirus, formally named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was made in Wuhan City, Hubei Province, China. In subsequent weeks, the coronavirus's rapid spread garnered increasing media coverage and public attention. Press coverage further heightened on January 21, 2020 when the CDC activated its Emergency Operations Center and the WHO began publishing daily

situation reports. Subsequent travel limitations, large-scale quarantine of Chinese residents, and numerous international index cases generated significant interest by the general public [9]. However, there is limited insight into the main topics discussed and the sentiment of the general public over time.

We postulate that analysis of the content and sentiments expressed over time on Twitter in the early stages of the coronavirus disease 2019 (COVID-19) pandemic can aid understanding of the effect of the outbreak on the emotions, beliefs, and thoughts of the general public. Such understanding would enable large-scale opportunities for education and appropriate information dissemination about public health recommendations.

METHODS

Data Collection

From January 14 to 28, 2020, a random sample of tweets in the English language was extracted using Twitter's application programming interface (API) and its advanced search tool (https:// twitter.com/search-advanced), which generates a relevant subset of tweets [10] that does not include any retweets. The dates were chosen to include 1 week of data before and after the activation of the Emergency Operations Center by the Centers for Disease Control and Prevention [11] and the release of the first WHO situation report [12]. Hashtags used for identification of COVID-19-related tweets included *#2019nCoV*, *#coronavirus*, *#nCoV2019*, *#wuhancoronavirus*, and *#wuhanvirus* (*COVID-19* and SARS-CoV-2 were not coined until February 19, 2020) based on the top trending hashtags related to the COVID-19 outbreak during the study period. Nineteen variables were extracted from tweets, 10 of which were used in our analysis: tweet text, time stamp, if the tweet had a reply, if the tweet was a reply, if the tweet was a retweet (which does not include quoted tweets), if the tweet included an image, if the tweet included a link, number of tweet likes, number of retweets, and number of replies.

Data Processing, Transformation, and Exploration

We performed all data processing and analysis using Python software, version 3.6.1 (Python Software Foundation) and RStudio version 1.2.1335 (R Foundation for Statistical Computing). We compared the COVID-19-related tweets per hour with the number of newly confirmed cases worldwide over each 24-hour period and completed descriptive statistics for the collected variables. To analyze tweets, we extracted the plain text from the original message. For all but the sentiment analysis, we removed commonly used words that are of little analytic value (eg, "for," "the," "is"), converted text to lowercase, and changed words to their root forms (eg, "viruses" to "virus" or "went" to "go"). We extracted 1-word and 2-word terms from tweets. We removed terms present in less than 5 tweets and 2 terms present in greater than 10% of tweets ("case" and "people") decreasing the dictionary of terms from 626 614 to 38 823.

Using a word cloud, we visualized the top 300 words with larger font size representing greater frequency. We used a subset of keywords to identify tweets related to 3 common infection prevention and control (IPC) strategies as well as vaccination. Appendix Section A1 details the keywords used. We analyzed the incidence of these tweets over time and manually reviewed a random 10% subset to validate content, evaluate narratives present, and explore examples of misinformation.

Sentiment Analysis

Sentiment polarity describes emotions that refer to the intrinsic attractiveness or aversiveness of a subject such as events, objects, or situations [13]. We analyzed the sentiment polarity of tweets separately using 4 commonly used methods through the *Syuzhet* R package [14]. Because each method uses a different scale, we normalized scores to detect the polarity of tweets as positive, negative, or neutral. For the emotion analysis, we used recurrent neural networks to label a primary emotion for a document according to a previously established emotional classification system (ie, anger, disgust, fear, joy, sadness, or surprise) [15]. We trended the findings by visualizing the daily number of tweets labeled with each sentiment polarity and each emotion over the 2-week period and comparing their rate of change by tweets per day.

Topic Modeling

A Latent Dirichlet Allocation (LDA) [16] model (*gensim* Python package [17]) automatically generates topics from observations (in our case, from tweets) and groups similar observations to 1 or more of these topics using the distribution of words. We iteratively trained multiple LDA models using different numbers of topics to maximize a topic coherence score (which measures the degree of semantic similarity between high-scoring words in the topic). Selecting the highest coherence score resulted in the use of the LDA model with 10 topics. Adhering to convention, we presented the top 15 terms (a common number of terms used in analyzing topics in LDA models) that contributed to each topic group and manually labeled a theme for each topic. We then visualized the topic model using a t-distributed Stochastic Neighbor Embedding (t-SNE) graph [18], which embeds high-dimensional data (ie, 10 dimensions given 10 topics) into a graphable 2-dimensional space where similar tweets are grouped together. We created an interactive visualization of the t-SNE to qualitatively evaluate the change in topics over time.

RESULTS

Tweet Frequency

A total of 126 049 tweets from 53 196 unique users were collected during the study period (Appendix Table A2). Of these

tweets, 123 407 had unique text (ie, text that was not duplicated in any other tweet in the dataset); there were no retweets in the sample. The most prevalent identification hashtag found was *#coronavirus* followed by *#wuhancoronavirus* present in 82% and 13% of tweets, respectively. The collected tweets accumulated 114 635 replies, 1 248 118 retweets, and 1 680 253 likes. In the first week of our analysis, the number of COVID-19-related tweets remained stable with less than 100 tweets per hour. The number of tweets per hour increased on January 20, 2020 and reached as many as 250 tweets per hour by January 21, 2020 and continued to grow with a peak of over 1700 tweets per hour by January 28, 2020. This trend closely tracked the number of newly confirmed COVID-19 cases in the study period (Figure 1).

Common Expressions

Collected tweets contained 2 877 816 words and 15 955 720 characters. The most common word in our analysis was "outbreak," numbering 11 549 times (Figure 2). The other top 15 most commonly used words and their frequency in descending order were as follows: "spread" (11 290), "health" (9734), "confirm" (6897), "death" (5819), "city" (5662), "report" (5662), "first" (5431), "world" (5244), "travel" (5049), "hospital" (4405), "infect" (4388), "SARS" (4133), "mask" (3996), "patient" (3981), and "country" (3885).

Infection, Prevention, and Control

Before January 20, 2020, our analysis showed a very small percentage of tweets related to IPC followed by a steady increase
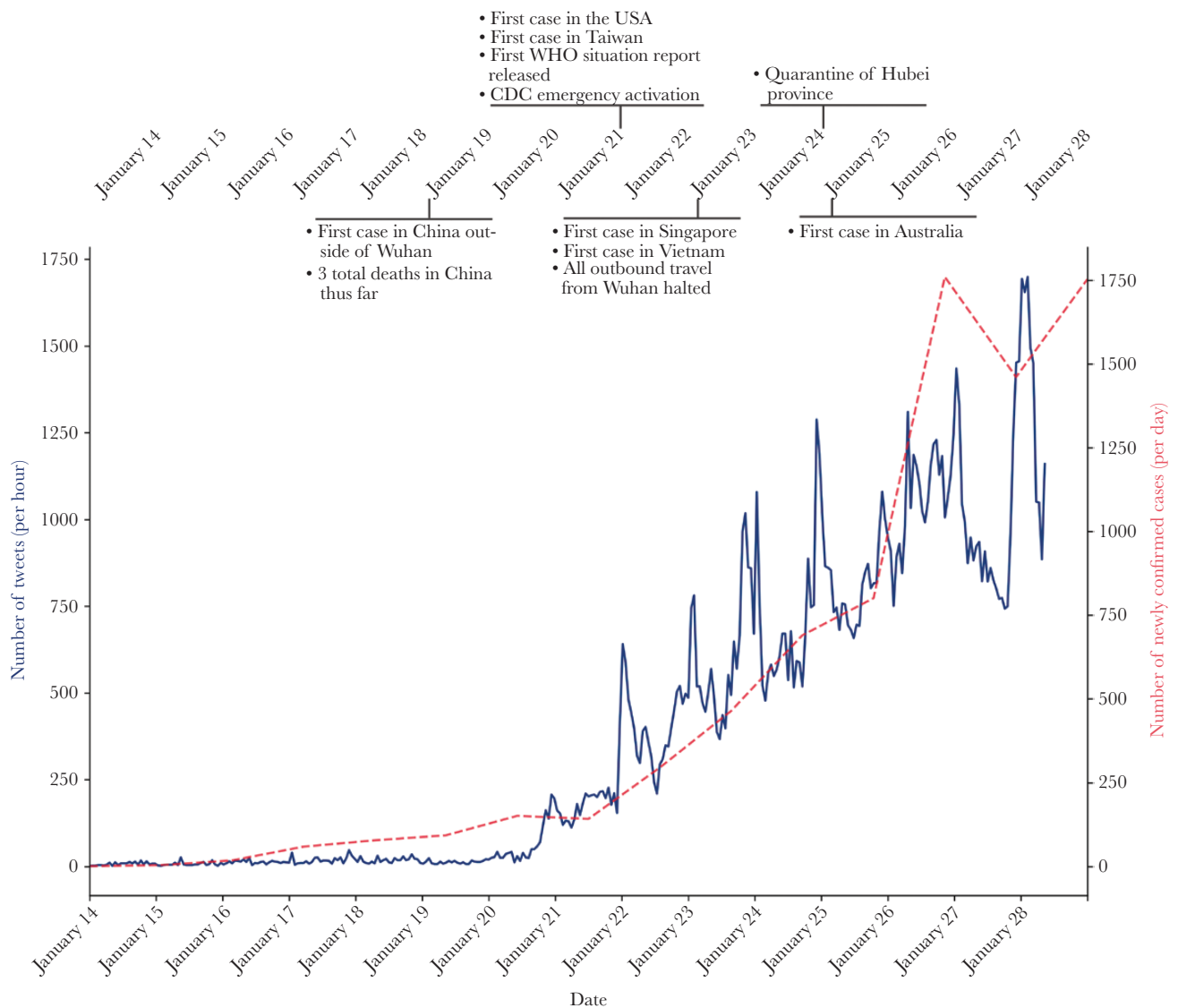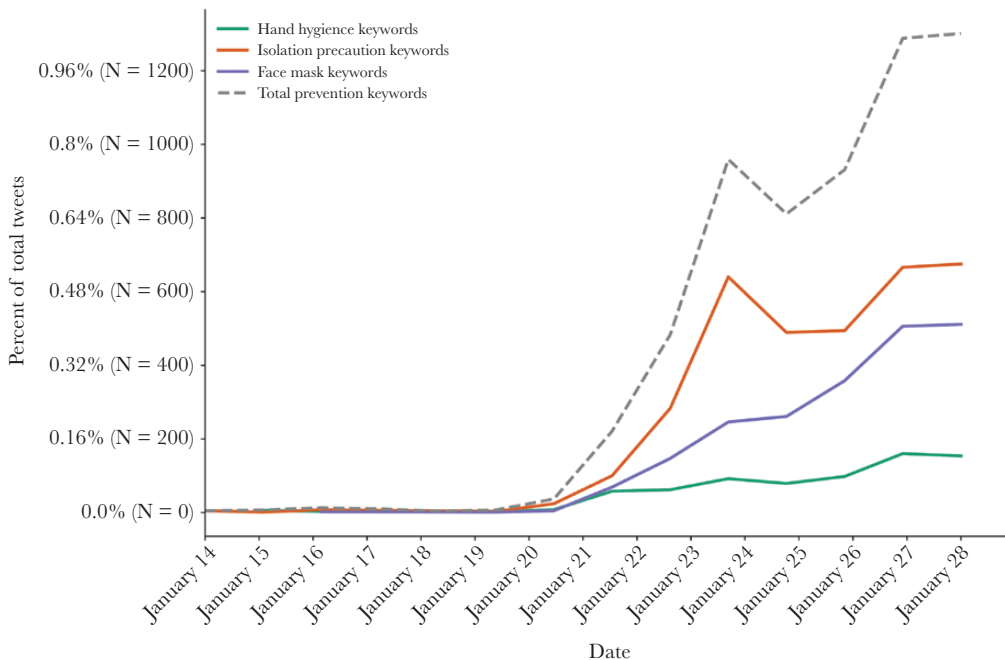

Figure 1. Number of coronavirus disease 2019 (COVID-19)-related tweets (left y-axis) and number of newly confirmed coronavirus cases (right y-axis) over time. CDC, Centers for Disease Control and Prevention; WHO, World Health Organization.
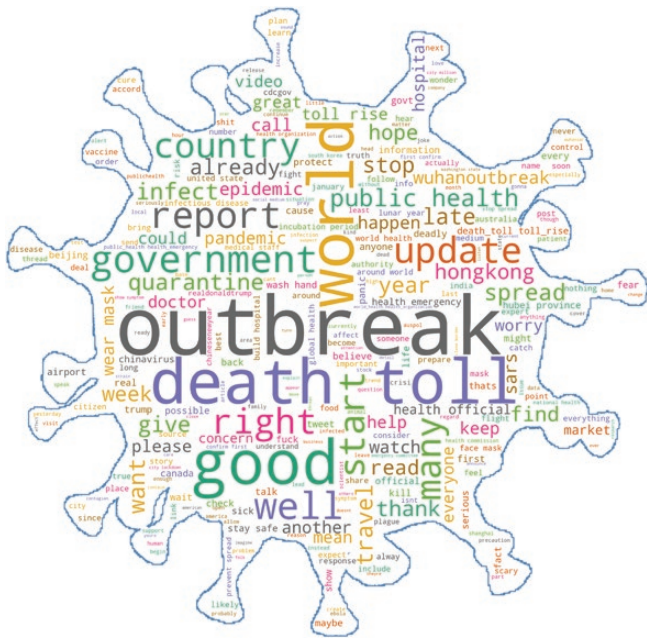
Figure 2. Word cloud showing the top 300 words used in tweets related to coronavirus disease 2019 (COVID-19).

starting January 21, 2020 (Figure 3). Isolation-related tweets were the most prevalent followed by mask and hand hygiene. Coinciding with the quarantine of the Hubei province, isolation-related tweets disproportionately increased on January 24, 2020. All IPC subgroups increased over time but their ranking did not change. The IPC-related content was present in 4.8% of tweets. Discuss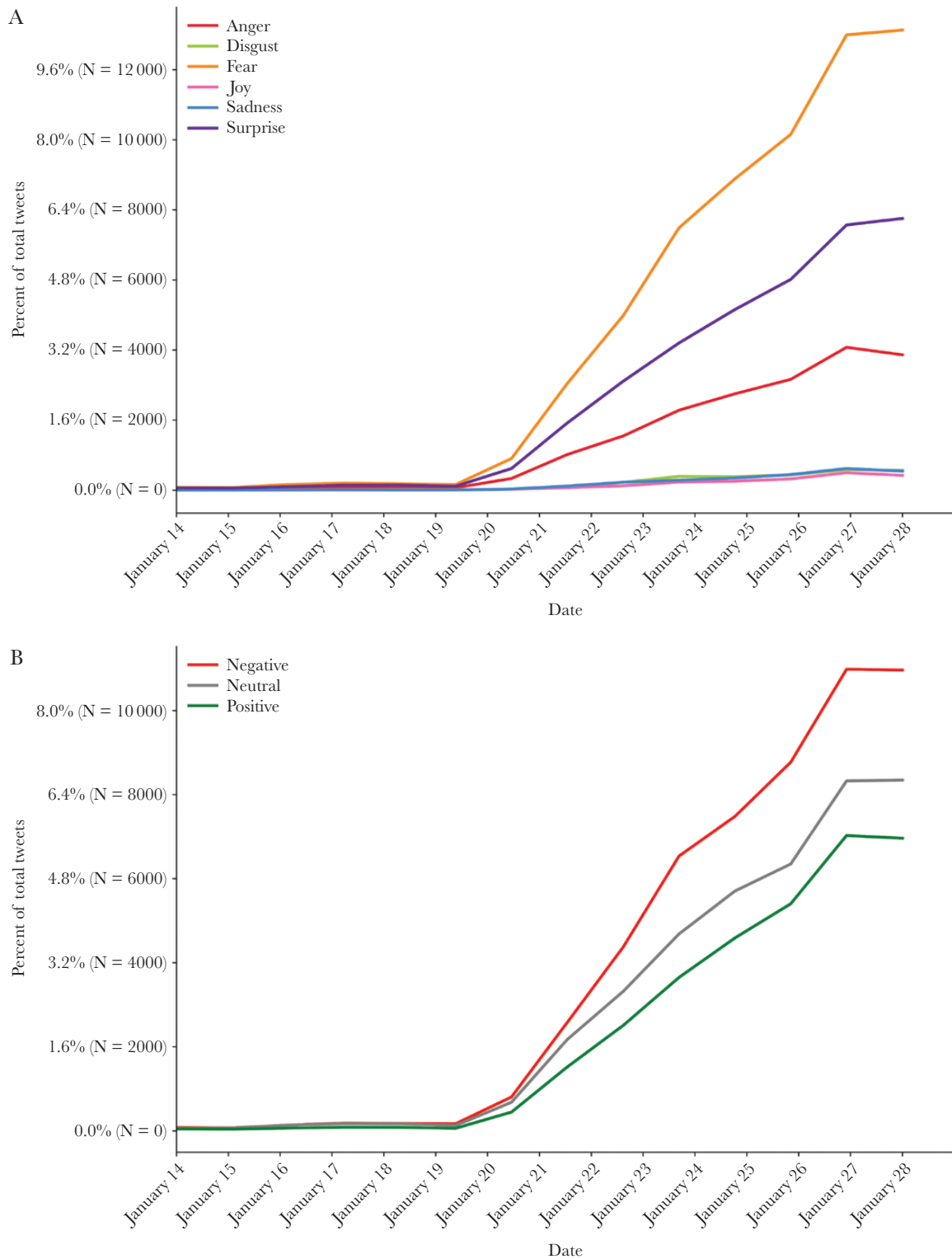ions of prevention techniques, shortage of protective gear, dissemination of health information, and large-scale quarantine were most common. Tweets with reference to vaccinations were found in 1.2% of total tweets and increased at a slower rate than IPC-related tweets overall. The most prevalent vaccine-related tweets were about vaccine availability, vaccine development, and advocacy to receive the influenza vaccine.

Sentiment Polarity and Emotions

Fear was the most common emotion expressed in 49.5% of all tweets with topics ranging from fear of infection, death, and inability to travel as well as emotional distress and fear regarding the effect on the economy and politics. [Examples: "*Coronavirus: Virus fears trigger Shanghai face mask shortage*" and "*Oil falls below $60 as China coronavirus fears accelerate*"] Surprise was the second most common emotion present in 29.3% of tweets. [Example: "*The Wuhan virus is more critical than expected! Don't forget to wear [a] face mask(surgical mask)!*"] Anger followed and included themes of inadequate governmental reactions, isolation and quarantine, lack of supplies, and lack of information. [Examples: "*Wuhan coronavirus: Hong Kong police, protesters clash as anger erupts over proposal to use housing block as quarantine site*" and "*11 million city on a lockdown!!!*"] The least common predominant emotions found in tweets were sadness, joy, and disgust (Figure 4A). More popular tweets contained less fear; 51.1% (n = 37 095) of non-retweeted tweets expressed fear compared with 41.3% (n = 49) of the top 0.1% retweeted tweets (Table 1). We analyzed tweets for positive, neutral, or negative polarity. Tweets with a negative sentiment polarity were more common than neutral and positive tweets

Figure 3. Daily number of tweets related to infection prevention and its subgroups of isolation/quarantine, masks, and hand hygiene.

A

B

Figure 4. Analysis of (A) tweet emotions (anger, disgust, fear, joy, sadness and surprise) and (B) sentiment polarity over time.

and increased at a faster rate over time (Figure 4B). Only the top 0.1% most retweeted tweets had an average neutral sentiment (median 0, interquartile range [IQR] −0.5 to 0.5). More sample tweets are included in Appendix Figure A2.

Topic Modeling

Topic modeling identified 10 themes that are recorded in Figure 5A. Keywords are listed in order of weight in forming the abstract topics found within the text. A tweet may include

Table 1. Comparing Sentiment Polarity, Emotion and Predominant Topics Among the Most and Least Retweeted Tweets[a]

Subset of (re)tweets	Polarity, Median (IQR)	Most Predominant Emotion		Most Predominant Topic		2nd Most Predominant Topic		3rd Most Predominant Topic	
		Emotion	N (%)	Topic	N (%)	Topic	N (%)	Topic	N (%)
Complete (n = 126 049)	−0.25 (−0.75 to 0.50)	Fear	62 424 (49.5)	Economic and Political Impact	20 385 (16.5)	Government Response	16 038 (13.0)	Outbreak/Pandemic	15 847 (12.8)
Zero retweets (n = 72 615)	0 (−0.75 to 0.5)	Fear	37 095 (51.1)	Economic and Political Impact	13 784 (19.4)	Government Response	9967 (14.1)	Outbreak/Pandemic	9221 (13.0)
Top 10% retweets (n = 12 604)	−0.25 (−0.75 to 0.50)	Fear	5506 (43.7)	Quarantine Efforts	1695 (13.4)	Outbreak/Pandemic	1552 (12.3)	Prevention	1498 (11.9)
Top 1% retweets (n = 1260)	−0.25 (−0.75 to 0.50)	Fear	533 (42.3)	Quarantine Efforts	168 (13.3)	Prevention	163 (12.9)	Outbreak/Pandemic	147 (11.7)
Top 0.1% retweets (n = 126)	0 (−0.50 to 0.50)	Fear	49 (41.3)	Quarantine Efforts	19 (15.1)	Healthcare Provision	19 (15.1)	Index Cases by Country	15 (11.9)

Abbreviations: IQR, interquartile range.

[a]Sentiment is shown as median and IQR. Emotion and the top 3 most predominant topics are shown as total number and percentage of tweets.

multiple topics, but it typically has 1 predominant topic. The most common predominant topic was the economic and political impact, followed by government response to the virus, then discussion of the outbreak and its development and transmission. The least common topics included index cases, the public health response, and healthcare provision. Other topics included the number of cases and death as well as prevention and large-scale quarantine. An interactive visualization of tweet themes showing their development by day is available at https://ssaleh2.github.io/Early_2019nCoV_Twitter_Analysis/; hovering over a node will show the tweet text and the day it was posted (please note the figure is slow to load and the slider on top allows navigation through time). Figure 5B shows 3 screen shots from the visualization. Major themes clustered in the center while more obscure tweets displayed in the periphery. Because tweets may include multiple topics, there is visible crossover between topic clusters in the visualization. Topic clusters that included themes of outbreak and its transmission, public health risk, and index cases were discussed from the start of the study period, whereas discussion of quarantine effects, economic and political impact, and government response increased significantly in the second week of the study period.

When focusing on the top 10%, 1%, and 0.1% most retweeted tweets, discussion of quarantine efforts was the most predominant topic (Table 1). Outbreak transmission as well as prevention were the next most common topics in the top 10% and 1% of tweets. In the top 0.1% of tweets, healthcare provision and index cases by country were the next most common topics.

DISCUSSION

In this study, we demonstrate a persistent increase in overall Twitter activity as well as tweets with negative sentiment and emotions for the COVID-19 outbreak from January 21, 2020 onward. The frequency of tweets paralleled the number of infected individuals worldwide during the early stages of the COVID-19 outbreak. Tweets predominantly showed negative sentiment and were linked to emotions of fear primarily, as well as surprise and anger. We identified examples of tweets with misinformation, but tweets were also significantly used to disseminate valuable public health information, especially in the more popular retweeted tweets. These data may help medical experts and public health officials to identify types of communication and messaging that may allay emotion and decrease misinformation.

Emotions have been shown to alter how we think, decide, and solve problems especially in highly charged situations of outbreaks [19]. Furthermore, "[p]atients' perception [...] of our health care system [...] informs, and is, their reality" [20]. For public health officials, governments, and healthcare industry leaders, understanding public sentiment and reaction to infectious outbreaks is crucial to predict utilization of healthcare resources and compliance with public health and infection prevention measures. Using the Streaming or PowerTrack API [21], Twitter allows access to the thoughts and emotions of millions of users and permits efficient and real-time analysis of these sentiments on important healthcare topics like the ongoing COVID-19 outbreak.

Surveillance programs for emerging and highly dangerous infections are difficult and labor intensive [22]. Leveraging the knowledge of the crowds by analyzing social media posts offers a simple and, in the case of the COVID-19 outbreak, a realistic view of the extent of the public health emergency. Despite collecting only tweets in English, the number of daily tweets paralleled the number of newly diagnosed cases even though most of these early cases were in China. The progression of fear and negative sentiment as well as the changes of topics discussed over time provided a granular view of early developing public discourse. Twitter may serve as a crucial culture medium for the growth and spread of public perception about global infectious outbreaks such as COVID-19.

A

Possibel topic label	Words contributing to topic model
Economic and political impact	market, stock, week, Trump, fear, watch, year, video, good, today, start, keep, spread, medium, even
Government response	world, spread, stop, outbreak, right, country, health, government, many, hope, good, happen, control, information, important
Outbreak/pandemic	SARS, outbreak, disease, pandemic, human, kill, could, animal, vaccine, spread, world, cause, Ebola, snake, global
Quarantine efforts	city, spread, million, travel, outbreak, safe, quarantine, lockdown, stay, close, province, Hubei, leave, border, prevent
Prevention	mask, symptom, health, Hong, Kong, tesk, Hong Kong, wear, patient, spread, show, wear mask, incubation, student, face
Increasing cases and death	death, confirm, hosiptal, number, toll, death toll, report, rise, infect, patient, spread, confirmed, total, toll rise, official
Media/news coverage	outbreak, update, infect, read, late, spread, person, thread, real, follow, good, contagion, global, plague, informaiton
Healthcare provision	hospital, life, doctor, risk, outbreak, government, please, medical, save, help, wuhanoutbreak, staff, protect, pray, nurse
Public health risk/travel	health, emergency, public, airport, public health, international, outbreak, global, flight, declare, screen, measure, passenger, world, health emergency
Index cases by country	first, confirm, report, States, Japan, Thailand, Australia, France, outbreak, Taiwan, Singapore, Korea, country, United, Canada

B

January 14   January 20   January 27
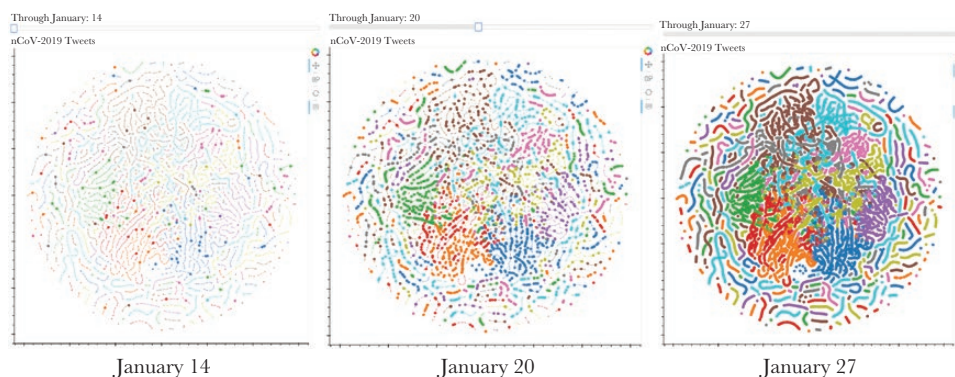
Figure 5. (A) The 15 terms (in order of weighting) that contributed to each abstract topic with their potential theme labels. The topics are ordered by frequency. Colors for each topic correspond to those in B. Topic labels were assigned by the authors. (B) A t-distributed Stochastic Neighbor Embedding (t-SNE) graph (17) (which embeds high-dimensional data into a 2-dimensional space where similar tweets are grouped together) that visualizes the topics in A as labeled by color and how they change over time. The full interactive visualization is available at https://ssaleh2.github.io/Early_2019nCoV_Twitter_Analysis/; please note the visualization is slow to load. Each node represents an individual tweet, and only tweets posted through the day highlighted on the slider are shown in the foreground, whereas all tweets in the study period are shown in the background. Hovering over a node will show the tweet text and the day it was posted. Depicted here are 3 screenshots for January 14, 2020 (day 0), January 20, 2020 (day 6), and January 27, 2020 (day 13).

Twitter is the most popular social media platform for healthcare communication; however, skepticism of its utility has been long discussed. Opponents often cite misinformation and the inability to process high volumes of information [23]. We found evidence of misinformation and hyperbole in tweets and reported online (Examples: "*People are literally dying on the streets of China [...]*," "*The new fad disease "coronavirus" is sweeping headlines. Funny enough, there was a patent for the coronavirus (sic) was filed in 2015 and granted in 2018,*" and "*Tesla Models S and X hospital grade HEPA filters may help prevent coronavirus infection*"). More sample tweets are available in Appendix Figure A2. Social media companies such as Facebook, Google, and Twitter have taken on the responsibility of acting as stewards of information related to COVID-19 by removing

false information and redirecting web traffic to reputable websites [24]. The account of the user, who tweeted the misleading patent information above was subsequently suspended [25]. Twitter Singapore adjusted their search prompt to show links to authoritative health sources such as the WHO and Ministry of Health for the COVID-19 outbreak [26]. Furthermore, it is important to point out that scientists and government officials also contributed to the dissemination of false information during this outbreak. A description of transmission in a prominent journal falsely reported that an asymptomatic person infected 4 others with coronavirus [27]. Researchers failed to interview the index case, who later reported that she had been symptomatic [28]. A since withdrawn scientific article falsely claimed that SARS-CoV-2 has 4 pieces of sequence in its genetic code not found in other coronaviruses and speculated that the virus could be genetically engineered [29]. The Chinese State media disseminated a fake photo of a newly erected hospital [30].

Despite evidence of misinformation, the most retweeted tweets ("viral tweets") were focused on topics to help disseminate knowledge of quarantine efforts, prevention, and information about the outbreak's spread. Crowdsourcing has been shown to be an enormously powerful and expedient way of achieving educational tasks [31]. The desire of the crowd to use a tool like Twitter to obtain and disseminate information offers the opportunity to change the narrative and educate millions of people. Since the outbreak started, the WHO has educated the public with a steady stream of tweets [32]. Some tweets analyzed were related to infection prevention measures (handwashing, mask wearing, self-isolation), but these were still the minority, representing less than 5% of tweets.

From a public health perspective, the ability to analyze Twitter feeds in real-time (using the Twitter Streaming or PowerTrack API) and the potential to individually target segments of the population with high-impact messages based on their information needs and sentiment could be an extremely powerful tool, potentially more effective than any other communication medium. To date, bots (autonomous programs able to interact with computer systems or users) have been used on Twitter for advertising or to promulgate malicious or false content [33, 34]. However, public health and governmental organizations such as the WHO or the CDC should invest in this new technology. Deploying autonomous tools that identify tweets, for example, by users who are scared to contract COVID-19, could be used to send individually targeted messages that provide reassurance and education on preventive measures such as handwashing and self-quarantine. Tailoring automatic responses to the sentiments and content of tweets has the potential to engage more Twitter users on public health topics and to redirect the discussion to useful, accurate information.

This study had several limitations. First, we used a noncomprehensive list of hashtags that was limited by a subset of trending hashtags at the time and the imagination of the authors. We may have missed alternative terminology or misspellings and may have introduced some selection bias in the tweets we analyzed. For example, #wuhanoutbreak was not included, but it arose as a weighted term in our topic modeling. In contrast, #coronavirus may have identified tweets related to other infections such as SARS. Second, despite the large number of tweets analyzed (>126 K), we collected and analyzed only a relevant subset of all tweets, which introduces some selection bias. Third, we targeted tweets in the English language; thus, our conclusions may not be generalizable to other countries where English is not the predominant language. Therefore, this study does not likely inform perception in China, where the majority of cases were in the early stages of the outbreak. Finally, we recognize that ascribing topic themes based on a subset of weighted terms has opportunity for labeling bias. To mitigate that, 2 authors designed the topic model and a separate set of authors labeled the topic themes.

CONCLUSIONS

We were able to show that the frequency of tweets paralleled the number of newly infected individuals for the early stages of the COVID-19 outbreak. Tweets predominantly showed negative sentiment and were linked to emotions of fear primarily, as well as surprise and anger. Although tweets with misinformation were present, tweets were also significantly used to disseminate valuable public health information, especially in the more popular retweeted tweets. Twitter offers novel opportunities to public health and governmental agencies to not only measure outbreaks, but also to target messages of a public health nature based on user interest and emotion.

References

1. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control **2010**; 38:182–8.

2. Chorianopoulos K, Talvis K. Flutrack.org: open-source and linked data for epidemiology. Health Informatics J **2016**; 22:962–74.

3. Signorini A, Segre AM, Polgreen PM. 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 **2011**; 6:e19467.

4. Househ M. Communicating Ebola through social media and electronic news media outlets: a cross-sectional study. Health Informatics J **2016**; 22:470–8.

5. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control **2015**; 43:563–71.

6. Shin SY, Seo DW, An J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci Rep **2016**; 6:32920.

7. Daughton AR, Paul MJ. Identifying protective health behaviors on Twitter: observational study of travel advisories and Zika virus. J Med Internet Res **2019**; 21:e13090.

8. Stefanidis A, Vraga E, Lamprianidis G, et al. Zika in Twitter: temporal variations of locations, actors, and concepts. JMIR Public Health Surveill **2017**; 3:e22.

9. Curtis Tate, Julia Thompson. State Dept. issues highest advisory: "Do not travel to China" amid coronavirus outbreak. *USA Today*. **2020**. Available at: https://www.usatoday.com/story/travel/news/2020/01/30/coronavirus-united-extends-china-flight-cancellations-march-28/4615127002/. Accessed 5 February 2020.

10. Search Tweets. Standard search API. Available at: https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets. Accessed 4 February 2020.

11. Centers for Disease Control and Prevention. First Travel-related Case of 2019 Novel Coronavirus Detected in United States. Available at: https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html. Accessed 5 February 2020.

12. World Health Organization. Novel Coronavirus (2019-nCoV) SITUATION REPORT—1. **2020**. Available at: https://docs.google.com/viewer?url=https%3A%2F%2Fwww.who.int%2Fdocs%2Fdefault-source%2Fcoronaviruse%2Fsituation-reports%2F20200121-sitrep-1-2019-ncov.pdf%3Fsfvrsn%3D20a99c10_4. Accessed 5 February 2020.

13. Frijda NH. The Emotions. Cambridge, New York, Paris: Cambridge University Press; Editions de la Maison des sciences de l'homme; **1986**.

14. Jockers ML. Syuzhet: extract sentiment and plot arcs from text. **2015**. Available at: https://github.com/mjockers/syuzhet. Accessed 30 January 2020.

15. Colneric N, Demsar J. Emotion recognition on twitter: comparative study and training a unison model. IEEE Trans Affective Comput **2019**:1.

16. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res **2003**; 3:993–1022.

17. Rehurek R, Sojka P. parallelized Latent Dirichlet Allocation. gensim. Available at: https://radimrehurek.com/gensim/models/ldamulticore.html. Accessed 30 January 2020.

18. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learning Res **2008**; 9:2579–605.

19. Jung N, Wranke C, Hamburger K, Knauff M. How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. Front Psychol **2014**; 5:570.

20. Betancourt JR. Perception is reality, and reality drives perception: no time to celebrate yet. J Gen Intern Med **2018**; 33:241–2.

21. Twitter, Inc. Filter realtime Tweets. Available at: https://developer.twitter.com/en/docs/tweets/filter-realtime/overview. Accessed 4 February 2020.

22. Yarbrough MI, Ficken ME, Lehmann CU, et al. Respirator use in a hospital setting: establishing surveillance metrics. J Int Soc Respir Prot **2016**; 33:1–11.

23. Pershad Y, Hangge P, Albadawi H, Oklu R. Social medicine: twitter in healthcare. JCM **2018**; 7:121.

24. Romm T. Facebook, Google and Twitter scramble to stop misinformation about coronavirus. The Washington Post. Available at: https://www.washingtonpost.com/technology/2020/01/27/facebook-google-twitter-scramble-stop-misinformation-about-coronavirus/. Accessed 5 February 2020.

25. Jones J. Misinformation about the coronavirus is spreading online. *HuffPost*. Available at: https://www.huffpost.com/entry/misinformation-about-the-coronavirus-is-spreading-quickly-online_n_5e2f6937c5b68f86c8ccbfee. Accessed 5 February 2020.

26. Twitter Singapore. Twitter Singapore Tweet. Available at: https://twitter.com/TwitterSG?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor. Accessed 5 February 2020.

27. Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med **2020**; 382:970–1.

28. Kai Kupferschmidt. Study claiming new coronavirus can be transmitted by people without symptoms was flawed. Science. **2020**. Available at: https://www.sciencemag.org/news/2020/02/paper-non-symptomatic-patient-transmitting-coronavirus-wrong. Accessed 5 February 2020.

29. Pradhan P, Pandey AK, Mishra A, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. Evolutionary Biology Cited March 2, 2020. **2020**. Available at: https://www.biorxiv.org/content/10.1101/2020.01.30.927871v2. Accessed 2 February 2020.

30. Lytvynenko J. Chinese state media spread a false image of a hospital for coronavirus patients in Wuhan. *BuzzFeedNews*. **2020**. Available at: https://www.buzzfeednews.com/article/janelytvynenko/china-state-media-false-coronavirus-hospital-image. Accessed 5 February 2020.

31. Bow HC, Dattilo JR, Jonas AM, Lehmann CU. A crowdsourcing model for creating preclinical medical education study tools. Academic Med **2013**; 88:766–70.

32. World Health Organization (WHO). World Health Organization Tweet. **2020**. Available at: https://twitter.com/WHO?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor. Accessed 5 February 2020.

33. Bob Abeshouse. Troll factories, bots and fake news: Inside the Wild West of social media. *Al Jazeera*. **2018**. Available at: https://www.aljazeera.com/blogs/americas/2018/02/troll-factories-bots-fake-news-wild-west-social-media-180207061815575.html. Accessed 5 February 2020.

34. Alex Hern. Microtargeting, bots and hacking: will digital meddling really swing this election? *The Guardian*. **2019**. Available at: https://www.theguardian.com/commentisfree/2019/nov/12/microtargeting-bots-hacking-digital-election-online-interference. Accessed 5 February 2020.