JEM

Data-driven inferences of agency-level risk and response communication on COVID-19 through social media-based interactions

Md Ashraf Ahmed, PhD Candidate
Arif Mohaimin Sadri, PhD
M. Hadi Amini, PhD, DEng

ABSTRACT

Risk perception and risk averting behaviors of public agencies in the emergence and spread of COVID-19 can be retrieved through online social media (Twitter), and such interactions can be echoed in other information outlets. This study collected time-sensitive online social media data and analyzed patterns of health risk communication of public health and emergency agencies in the emergence and spread of novel coronavirus using data-driven methods. The major focus is toward understanding how policy-making agencies communicate risk and response information through social media during a pandemic and influence community response—ie, timing of lockdown, timing of reopening, etc.—and disease outbreak indicators—ie, number of confirmed cases and number of deaths. Twitter data of six major public organizations (1,000-4,500 tweets per organization) are collected from February 21, 2020 to June 6, 2020. Several machine learning algorithms, including dynamic topic model and sentiment analysis, are applied over time to identify the topic dynamics over the specific timeline of the pandemic. Organizations emphasized on various topics—eg, importance of wearing face mask, home quarantine, understanding the symptoms, social distancing and contact tracing, emerging community transmission, lack of personal protective equipment, COVID-19 testing and medical supplies, effect of tobacco, pandemic stress management, increasing hospitalization rate, upcoming hurricane season, use of convalescent plasma for COVID-19 treatment, maintaining hygiene, and the role of healthcare podcast in different timeline. The findings can benefit emergency management, policymakers, and public health agencies to identify targeted information dissemination policies for public with diverse needs based on how local, federal, and international agencies reacted to COVID-19.

Key words: COVID-19, Twitter, social media, public organizations, interactions, historical data, policy

INTRODUCTION

No person or place is resistant from disasters or losses resulting from an extreme event. Infectious disease, in addition to natural hazards, wildfires, acts of terrorism, civil instability, or financial disasters may all have far-reaching effects for the country and its communities.[1] One way to reduce the impact of disasters on the nation and its communities is by investing in resilience enhancement similar to infrastructure resilience.[2] Building disaster resilience culture and practice, however, is not simple or inexpensive. Bottom-up initiatives (communities' dedication to improving their resilience) are important as local conditions differ widely throughout the world. Communicating risks, connecting community networks,[3] and promoting resilience culture can aid local communities in making progress to increase their resilience. Throughout this regard, community coalitions of elected representatives from the public and private sectors, with ties and funding from federal and state governments, and with input from local residents, are becoming very relevant. Such coalitions

DOI:10.5055/jem.0589

may be charged to determine the exposure and vulnerability of the community to hazard, to inform and communicate risk, and to analyze and extend the capacity of the community to manage such risk. A truly robust coalition will have a clear leadership and governance system at its heart, and people with sufficient time, expertise, and commitment required to establish and sustain relationships among all the community stakeholders.[4]

Public agencies engagement in risk communication can lead to more effective decision-making and enhanced public feedback to the regulatory process. The primary goal of this study is to mine and analyze large-scale time-sensitive perishable crowd-sourced and social media data (rich spatiotemporal data) and reveal interaction patterns of health-risk communication and community responses of major public agencies in the emergence and spread of novel coronavirus (nCov) using data-driven methods and network science theories. The specific aims are twofold: (1) to document how public agencies interact and communicate health risk information through their online social networks during a major disease outbreak and (2) to examine how online social networks influence the current COVID-19 pandemic situation in terms of the change in daily number of cases and deaths in the United States. To achieve the goal and aims, this study has utilized time sensitive social media interactions of six major public agencies: World Health Organization (WHO), Center for Disease Control and Prevention (CDC), Federal Emergency Management Agency (FEMA), American Red Cross, New York City (NYC) Department of Health and Mental Hygiene, and Mount Sinai Health System, NYC with crowd-sourced information on COVID-19.

This study has four theoretical and methodological contributions as compared with the literature:

(1) It has developed methods to advance our understanding of how public agencies influence online, while communicating health risks and interacting in their respective communities as the disease continues to spread.

(2) It informed the literature on how information is exchanged among people who are socially connected online and exposed to health risk in such outbreaks of disease.

(3) This study used novel machine learning models to quantify the influence on such communication behavior.

(4) It captured the variability in risk communication strategies and risk averting behaviors adopted based on temporal correlations of risk and disease contagion.

The findings from this research will be useful to public health and emergency management agencies for tailoring effective information dissemination policies for diverse user groups based on their social network characteristics, activities, and interactions in response to similar public health hazards. The methodologies and implications of this research can be transferred in designing effective intervention policies to other natural and man-made disaster contexts, in which public health risks become major concerns.

LITERATURE REVIEW

The transmutability of the nCov disease in 2019 was tracked in Zhejiang, China, accounting for transmissions from imported cases. While Zhejiang is one of the worst 21 affected provinces, an interruption in the transmission of the disease, ie, an instant reproduction number < 1, was observed in early/mid-February 2020 following an early social-distancing response to the outbreak.[5] Zhang et al. developed data-driven, susceptible-exposed-infectious-quarantine-recovered models to simulate the COVID-19 outbreak with measures of social distancing and epicenter lockdown. Population migration data combined with officially recorded data were used to estimate model parameters and then to measure the daily exported infected individuals by estimating the daily infected ratio and the daily susceptible population size.[6]

In 2003, public health interventions were crucial in the prevention of the Severe Acute Respiratory Syndrome (SARS) epidemic. Community containment requires steps ranging from increasing social distancing to city-wide quarantine. Whether these steps (isolation, quarantine, and social distancing) will be appropriate to monitor 2019-nCoV depends on resolving some unanswered issues.[7] Many countries only seek to achieve social distancing and hygiene measures when widespread transmission is apparent. It gives the virus a number of weeks to propagate with a higher basic reproductive number than if it had been in place before transmission was observed or widespread. Hence, preventive, low cost, enhanced hygiene, and social distance in the sense of imminent population transmission of the nCoV COVID-19 should be considered.[8]

Social media is considered as an accessible and widely used platform for efficient crisis communication in recent literature.[9-11] Traditional media is primarily intended for one-way communication, whereas social media allows two-way communication; hence, social media platforms are quite different and provide a wider range of data.[12,13] Sadri et al. explored the critical role of social media during crisis by facilitating communication and information dissemination to both evacuee and nonevacuee during Hurricane Sandy.[14] Zhang et al. envisage intelligent public disaster information and alert based on social media, which has three functions: (1) effectively and efficiently collecting disaster situational awareness information[15]; (2) promoting self-organized assistance activities; and (3) enabling emergency management agencies to hear from the public. The results of this analysis highlight the importance of such research fields: (1) a fine-grained social media catastrophe ontology with semantic interoperability, (2) trend knowledge network pattern and emerging prominent users, (3) fine-grained societal impact assessment due to infrastructure failures, and (4) best practices for social media use during disasters.[16]

Austin et al. examine how audiences seek social and traditional media information, and what factors influence media usage during crises. Using the model of social-mediated crisis communication, a review of crisis information and sources reveals that audiences use social media for insider information and check-in with family/friends during crises, and use traditional medias, such as television, radio, newspaper, etc., for increasing public awareness. Convenience, interaction, and personal feedback encourage the use of conventional and social media; usage of both discourages overload. Humor and beliefs toward social media uses discourage the use of social media, while legitimacy promotes the use of conventional media. The findings stressed the importance of the influence of third parties in crisis communication and the need to use traditional and social media in crisis response.[17]

Freimuth et al. describe the design, implementation, and assessment of a risk communication simulation during the first hours of a pandemic. The simulation design was focused on the communication of crisis and emergency risks principles upheld by the CDC, as well as the author's collective experience. Several local health district risk communicators in Georgia responded to a scenario where after returning from an international conference, every community in the state had teenagers infected with avian flu. The evaluation revealed that under the time pressures of a realistic and stressful event, local risk communicators had much greater difficulty following the principles of risk communication than they did in a tabletop workout. Strengths and weaknesses of local risk communicators' performance are identified in addition to the lessons learned on designing and implementing a simulation for risk management.[18]

Palen et al.[19] studied the rapid growth of social media in a number of disaster situations, exploring issues, such as citizen engagement, community-oriented computing, distributed problem solving, and digital volunteerism as modes of sociotechnical innovation, as well as issues of situational knowledge and truthfulness as opportunities and challenges emerging from the social media data deluge. This article also discusses the study that looks at integrating social media technologies and data into current emergency response work. Reflecting on the decade-old area of science, the authors warned of the danger that all "crisis" encounters can fail unintentionally without differentiation, which appears to happen because social media networks cross-cut all emergency situations. Through an effort to isolate what social media adds new, there is a tendency to fail to recognize how nontechnological influences on cultural sociobehavioral scales greatly affect the usage of social media itself.

Misinformation spreading in social media is also becoming an evolving concern. Monahan et al. mentioned that mass media play an important but often misunderstood role in the events of a catastrophe. Research has repeatedly shown that disaster-related media reporting appears to be riddled with disinformation and promotes misconceptions about race, social status, aggression, and crime. Studies have found that powerful media campaigns can support prevention efforts, strengthen early warning systems, facilitate orderly, prompt evacuation procedures, and help bring communities together in times of upheaval. Authors review research on the relationship between media and disaster to highlight the many ways media can positively impact disaster preparation and recovery, while also highlighting the many issues associated with disaster reporting. Future directions for media–disaster research are being discussed along with ways in which media staff and emergency response professionals can handle the media–disaster relationship more efficiently before, during, and after emergency incidents.[20]

Battur et al. detected twitter bot, which is a software that sends fake tweets automatically to users. Detecting bots is necessary to identify the fake users and to protect the genuine users from misinformation and malicious intents. This study proposes an approach to detect the twitter bots using several machine learning algorithms, such as Decision Tree, Multinomial Naïve Bayes, Random Forest, and Bag of Words. The algorithm with highest accuracy (Bag of Words) is used to test real time data.[21] Asr et al. stated that misinformation detection at the level of full news articles is a text classification problem, and reliably labeled data in this domain are rare. Previous work relied on news headlines, microblogs, tweets, and articles collected from so-called "reputable" and "suspicious" websites and labeled accordingly. Authors leveraged fact-checking websites to collect individually labeled news articles with regard to the veracity of their content.[22] Huang et al. proposed a systematic meta-analysis looked at 38 studies involving real responses to hurricane warnings and 11 studies with expected responses to hypothetical hurricane performed since 1991.[23]

METHODOLOGY

The focus of this study is to examine how public agencies' risk assessments, risk averting behaviors, and crisis communication patterns in online social network influence the pandemic situation using a data-driven solution that leverages Twitter data. Several machine learning algorithms have been used in this study to reveal different agencies' interaction patterns. First, a sentiment analysis over time for major public organizations had performed. Then, dynamic topic model (DTM) (based on Topic Model theory) was applied to identify the topics over the same time horizon that causes the change of sentiments.

Sentiment analysis is a method of natural language processing (NLP) task at many levels of granularity. Starting from a document level classification task, it has been deployed at the sentence level and more recently at the phrase level.[24] It is well recognized that twitter user-generated content with rich sentiment information should be utilized for many applications such as search engines and other information systems. While tweet level sentiment analysis results indeed provide very useful information, the overall or general sentiment tendency toward topics is more appealing in some scenarios. For example, different stakeholders, eg, advertising companies, are curious about how others feel about Apple's new product, "iPhone11," and it will offer great convenience for them if major opinions are collected from massive tweets.[25] To derive the sentiment scores, a text sentiment analysis model named VADER (valence aware dictionary for sentiment reasoning)[26] is used, which considers both the polarity (positive/negative) and the force (strength) of emotion. It calculates the positive, negative, and neutral sentiment score for each string (set of words) and reports the compound (compute by normalization) score directly. The sentimental study of VADER is based on a dictionary that maps lexical characteristics to emotion intensities, which are referred to as sentiment ratings. A text's sentiment score can be calculated by including the strength of each word in the text.[27]

A topic model is a statistical model to explore the abstract "topics" that happen in an assembly of information in machine learning and NLP. This was initially explored by David Blei according to the most common topic model named Latent Dirichlet Allocation (LDA). The idea behind LDA is that the set of texts reveal numerous topics. In topic models, first, the algorithm selects a topic, and then sample a set of words from the given topic. Clusters of similar terms are the themes or topics created by topic modeling technology. A topic model is a recurrently performed text-mining tool for the discovery of hidden semantic structures in a text body. Topic models can help us to organize and provide insights into understanding large collections of unstructured text bodies.[28]

The DTM includes a group of probabilistic time series model, which is used to observe the time evolution of topics in huge document collections. This group of models was proposed by David Blei and John Lafferty and is an extension to LDA that can handle chronological documents. In LDA, both the order and the words appear in a document, whereas words are still assumed to be interchangeable, but in a DTM, the order of the documents plays a key role. The method is to use state-space models to represent the topics on the natural parameters of the multinomial distributions. To perform approximate posterior inference over the latent topics, variation in approximations based on Kalman filters and nonparametric wavelet regression are developed. In addition to providing sequential, quantitative, and predictive models, DTM provides a qualitative window into the contents of a large document collection.

It is assumed that the data are divided by time intervals in a DTM, for example, by month. It is modeled as each slice's documents with a K component subject model, where slice t-related topics evolve from slice t-1-related topics. Let β_t denotes the V-vector of natural parameters for topic k in slice t for a K-component model with V terms. A multinomial distribution is usually represented by its mean parameterization. If we denote the mean parameter of a V-dimensional multinomial by π, the mapping $\beta_i = \log(\pi_i/\pi_V)$ of the ith element of the natural parameter is given. Dirichlet distributions are used in typical language modeling applications to model uncertainty over word distributions. The Dirichlet, however, is not conducive to sequential modeling. Alternatively, we chain the natural parameters of each β_t subject into a state-space model that evolves with Gaussian noise; the following equation shows the simplest version of such a model.

$$\beta_t \mid \beta_{t-1} \sim N(\beta_{t-1}, \sigma^2 I) \tag{1}$$

In LDA, the document-specific topic proportions θ are drawn from a Dirichlet distribution. In the DTM, a logistic normal is used with mean α to express uncertainty over proportions. The sequential structure between models is again captured with a simple dynamic model, which is expressed by the following equation:

$$\alpha_t \mid \alpha_{t-1} \sim N(\alpha_{t-1}, \sigma^2 I) \tag{2}$$

For simplicity, it did not model the dynamics of topic correlation, as it was done for static models by Blei and Lafferty.[29] By chaining together topics and topic proportion distributions, it has sequentially tied a collection of topic models.

DATA SOURCE AND DATA COLLECTION

Existing traditional datasets have limited capacity to adequately capture user risk communication strategies and analyze user concerns with such details and coverage. As such, social media datasets, enriched with user activity information, will be useful to capture user sentiments and concerns in real time and help early detection of the people exposed to health-risks in the vulnerable communities. Twitter, in particular, provides unique features to release their data through their application programing interface (API) and make it publicly available which could then be combined with other complementary information, eg, crowdsourced data, in particular over the timeline of COVID-19 crisis. For this study, Twitter-search APIs is used to collect and store public agencies' crisis interactions through social media outlets. NLP and machine learning techniques are adopted here to extract user concerns, response, and needs over time. However, Worldometer data (crowdsource data) is used for extracting the daily number of cases and deaths due to COVID-19 for the United States as it is one of the most reliable and trustworthy sources.[30] The main reason for considering these data is to compare the prominent discussions on Twitter along with the changes in number of cases and deaths due to COVID-19.

The goal of this study is to reveal some specific agencies' perspective during COVID-19 that exists in social media interactions. To achieve this goal, the first key data source we considered is the Twitter data. We were particularly interested in the tweets generated from major public health agencies (WHO, CDC, and NYC Department of Health), disaster management organizations (FEMA and American Red Cross), and hospitals (Mount Sinai Health System, NYC). Hence, historical tweets were collected during the COVID-19 pandemic of these six major agencies to perform the text-based analyses. All the twitter data were collected from February 21, 2020 to June 6, 2020 (around three and half months), which covers the initial surge and recession of COVID-19. As New York was the main hotspot of COVID-19 during this timeline, two specific organizations from New York (NYC Department of Health and Mount Sinai Health System, NYC) have been considered along with the other four major public organizations of the United States. The summary of the collected twitter data shows that WHO was the most active on Twitter than any other organizations. Twitter handles (username of the Twitter page) of the organizations have also provided in parenthesis with the number of tweets in the following list:

- WHO: 4,434 tweets (@WHO)
- CDC: 868 tweets (@CDCgov)
- FEMA: 1,996 tweets (@fema)
- American Red Cross: 809 tweets (@RedCross)
- NYC Department of Health and Mental Hygiene: 891 tweets (@nycHealthy)
- Mount Sinai Health System, NYC: 2,026 tweets (@MountSinaiNYC)

The following flowchart (Figure 1) is showing the methodology of the data analytics for this study.

ANALYSES AND RESULTS

To reveal the interaction patterns of the public organizations, several machine learning algorithms, including sentiment analysis and topic frequency, are applied over time. The graphical representations and

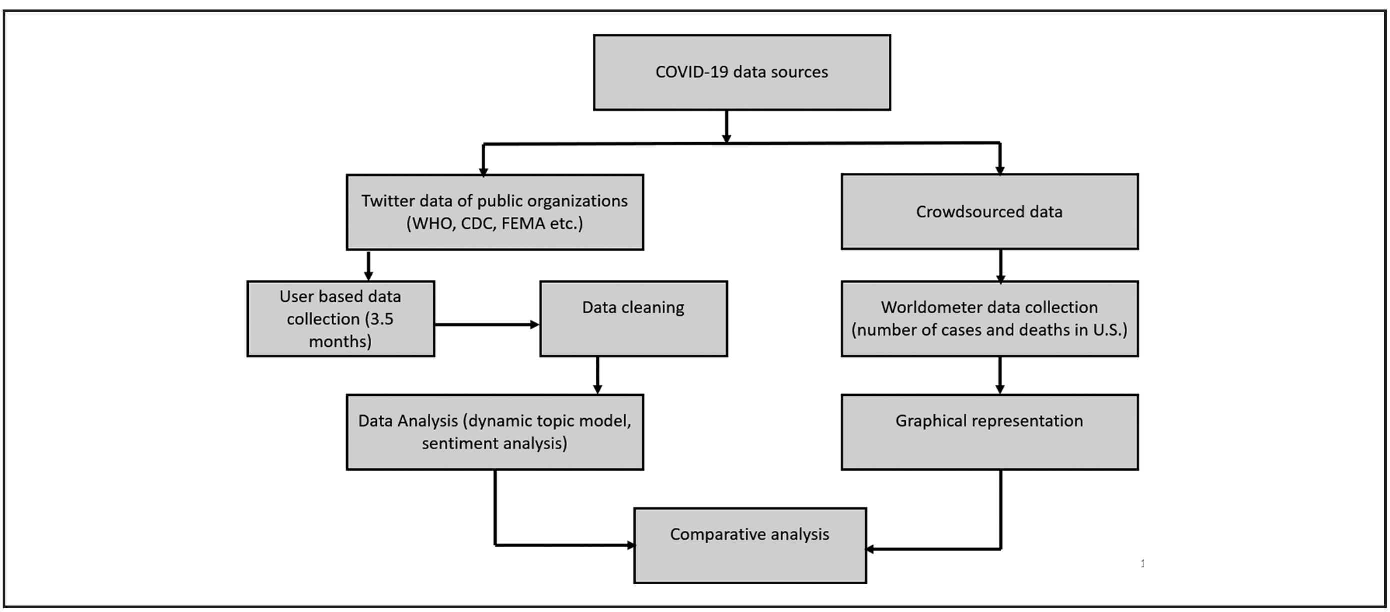

Figure 1. Flowchart of the methodology of the study.

the result interpretations of the tweets from six public organizations are listed later. Figures A1-A12 are shown in Appendix A.

WHO interactions

To understand the dynamics of the crisis communication pattern, 15 optimum number of topics were identified from static topic model analysis of WHO tweets. In Figure A1, the topic frequencies are represented along with the time (all the dates are in the year of 2020). Furthermore, in Figure A2, average sentiment scores of tweets are illustrated over the same timestamp. WHO emphasized on community transmission and the need of vaccine in March; shortage of healthcare workers was discussed (Figure A1) at the beginning of April; shortage of personal protective equipment (PPE), need of staying home, and importance of wearing masks became prominent from late April and in May 2020. However, requirement of government support got notified at the beginning of May, the number of cases (*COVID-19 in countries* topic) found increasing after 1.5 months' time interval. In Figure 2, the word probabilities over the timeline of each topic are visualized from which the change in social distancing phrase to physical distancing is observed in topic 10.

Several other topics, such as health concerns, importance of physical activities, responses from different countries, social measures, and use of tobacco, received attention also from WHO in twitter.

By comparing Figures A1 and A2 (rolling mean or 1-day running average), we observe that the tweet sentiment went to most negative, while WHO emphasized on alarming number of cases of COVID-19, lack of peoples' responses, shortage of PPE, and negative impact of using tobacco. Then, positive sentiments have found when WHO discussed about the need of government support, responses from people, development of vaccine, importance of home quarantine, and use of tobacco to develop the vaccine. Hence, the use of tobacco considered as negative topic in the beginning, but later it turned into positive topic.

CDC interactions

From CDC tweets, eight optimum number of topics are plotted over the three- and half-month timeline in Figure A3. CDC put importance on following their guideline consistently, increase in number of cases in whole March, hospitalization rates at the end of April which also found frequent after 1 month, pandemic stress in May rather than from the beginning, prevention of spread during the mid-April, contact tracing at the beginning of April, and risk of older people in March and May. In Figure 3, the word probabilities over the time of each topic are presented.

After comparing Figures A3 and A4, positive spikes in tweet sentiment have found when CDC put emphasize on following their guideline from the beginning and mostly on May, importance of contact tracing at the beginning of April, reducing the spread of the virus as well as the pandemic stress. However, negative sentiments are observed when elderly people at risk; increased number of cases and hospitalization rates became more frequent in twitter.

FEMA interactions

From around 2,000 tweets of FEMA, 10 optimum number of topics word probabilities over the timeline are visualized in Figure 4.

FEMA discussed about the importance of critical information in April, need of COVID-19 responses in mid-March and also after 1 month (mid-April), spreading of the virus at the end of April, and need of supplies and food in April and May. Hurricane season also got attention of FEMA during the pandemic. However, FEMA emphasized on the importance of COVID-19 testing from April to May, which is not mentioned by WHO and CDC. Also, lack of medical supply got notified from end of the March, April, and May, which show after each month, medical supply in the United States is needed.

By interpreting Figures A5 and A6 together, the extreme negative sentiments are observed when the need of medical supplies, lack of food, and the emergence of upcoming hurricane season along with the increased spreading of the virus are discussed. Then, positive sentiments have found while the importance of COVID-19 testing, availability of responses, fund, and spreading of critical information got more importance.

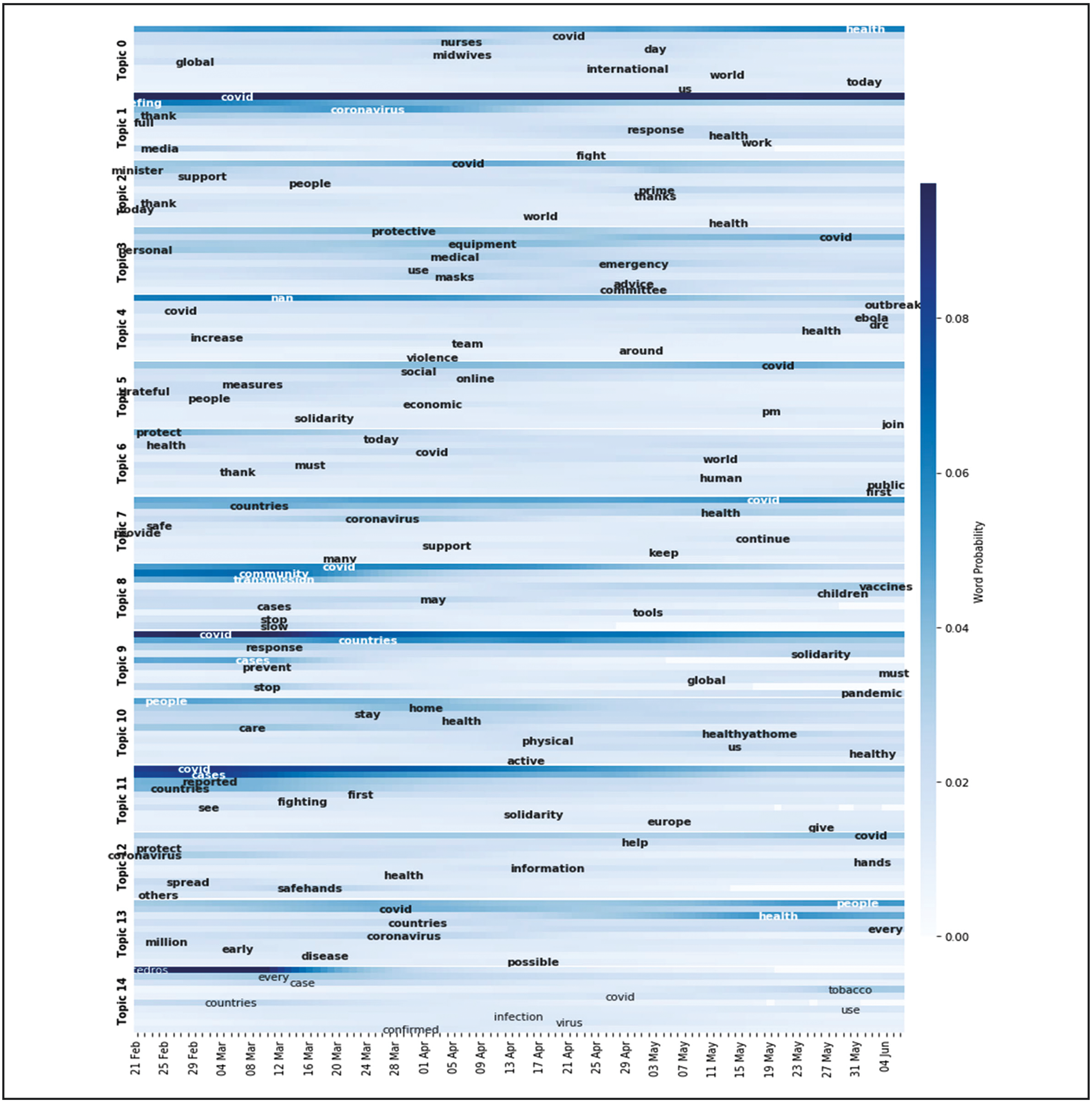

Figure 2. Word probability over time in topics from WHO tweets.

American Red Cross interactions

American Red Cross tweets showed eight optimum topics which are plotted over time in Figure 5.

America Red Cross emphasized on taking care of sick people due to COVID-19 from the mid-March and showed willingness to provide more help to the COVID-19-affected people in May frequently, blood donation and the eligibility requirements to donate blood from the US Food and Drug Administration (FDA) in March. However, they also focused on the

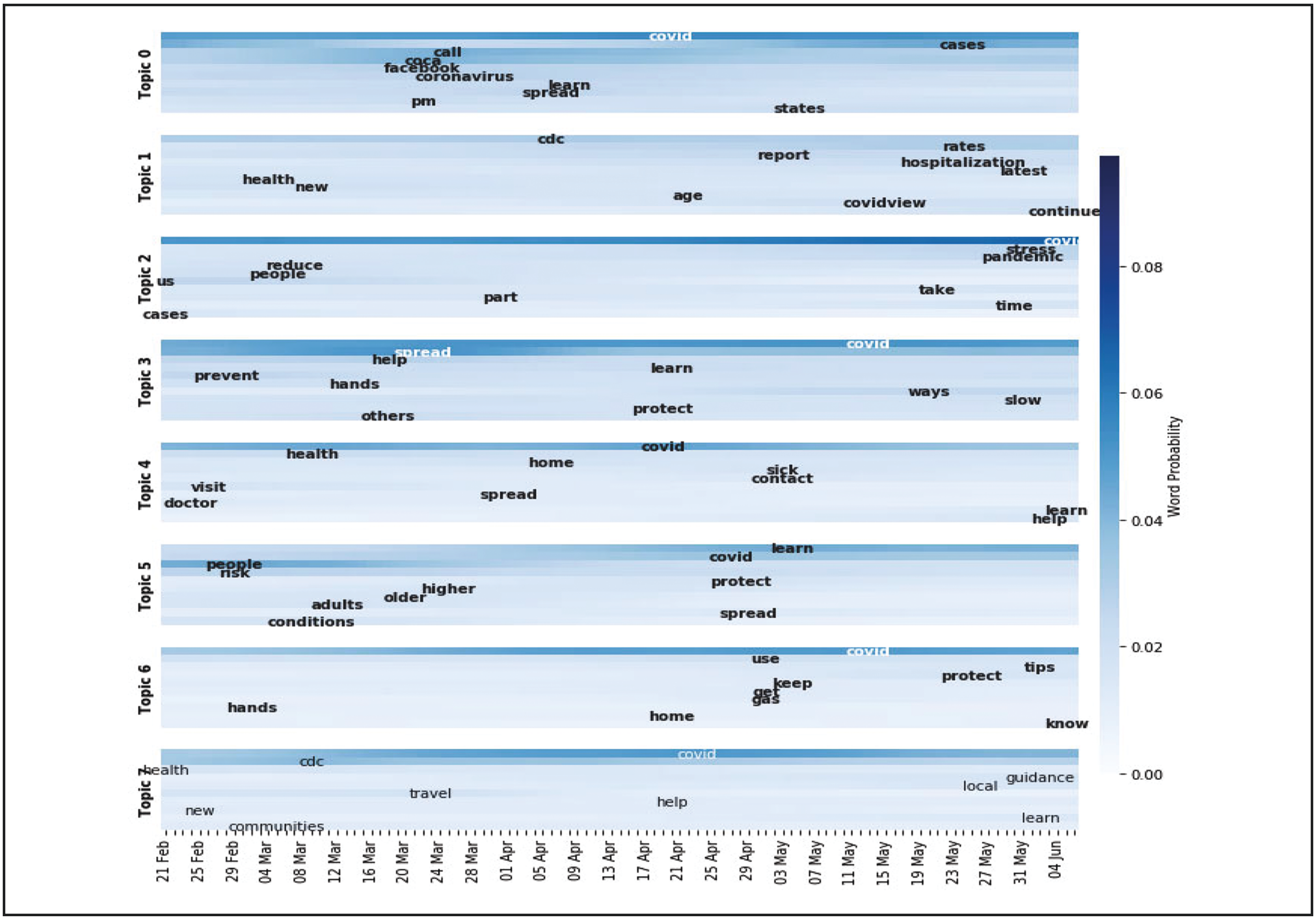

Figure 3. Word probability over time in topics from CDC tweets.

volunteer support consistently, importance of social distancing, and COVID-19 testing at the end of May. The most unique topic they mentioned is the importance of convalescent plasma donation to cure the COVID-19-infected patients throughout the timeline.

By comparing Figures A7 and A8, spikes in positive sentiments are found when American Red Cross put importance of taking care of sick people due to COVID-19, exhibited willingness to provide more help to the affected people, and discussed the importance of social distancing to slow down the spread of nCoV and the effectiveness of convalescent plasma to cure the COVID-19-infected patients. On the other hand, negative sentiments have observed while the eligibility requirements to donate blood from the FDA and the lack of volunteer support topics became more frequent in Twitter.

New York City Department of Health and Mental Hygiene interactions

From the NYC Department of Health tweets, eight optimum number of topics are visualized over the three- and half-month timeline in Figure 6.

They focused on pandemic stress reduction and the spreading of the nCoV consistently throughout the timeline due to the outbreak and were concerned about the health risk and COVID-19 responses at the beginning of March. However, they also emphasized on staying home and symptoms of COVID-19 at the end of March and provided health tips consistently on twitter during the mentioned timeline.

After comparing Figures A9 and A10, positive sentiments of tweets have been found when NYC Department of Health emphasized on the importance

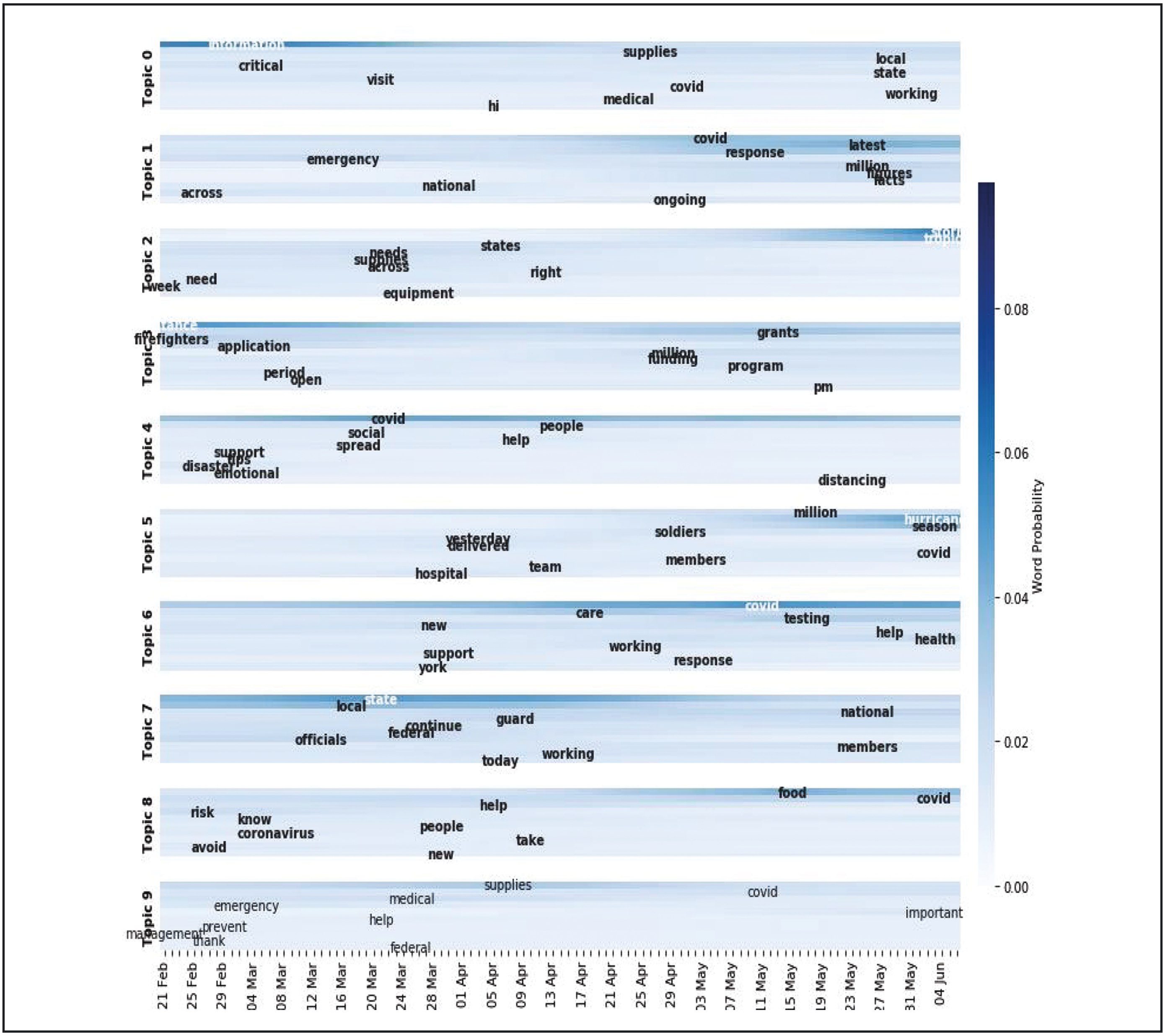

Figure 4. Word probability over time in topics from FEMA tweets.

of staying home to slow down the spread of coronavirus, specified the symptoms of the disease, provided the health tips for fighting the pandemic, and was concerned about the COVID-19 responses. However, negative sentiments are observed when the spread of nCoV, concerns about health risk, and stress induced due to the outbreak discussed more recurrently in twitter.

Mount Sinai Health System interactions

From more than 2,000 tweets of Mount Sinai Health System, 11 optimum number of topics frequency are visualized in Figure A11. Mount Sinai discussed about the need of frontline support from April, medical care and a surge of cancer patients in March, and the importance of contact tracing at the end of February. They also emphasized on

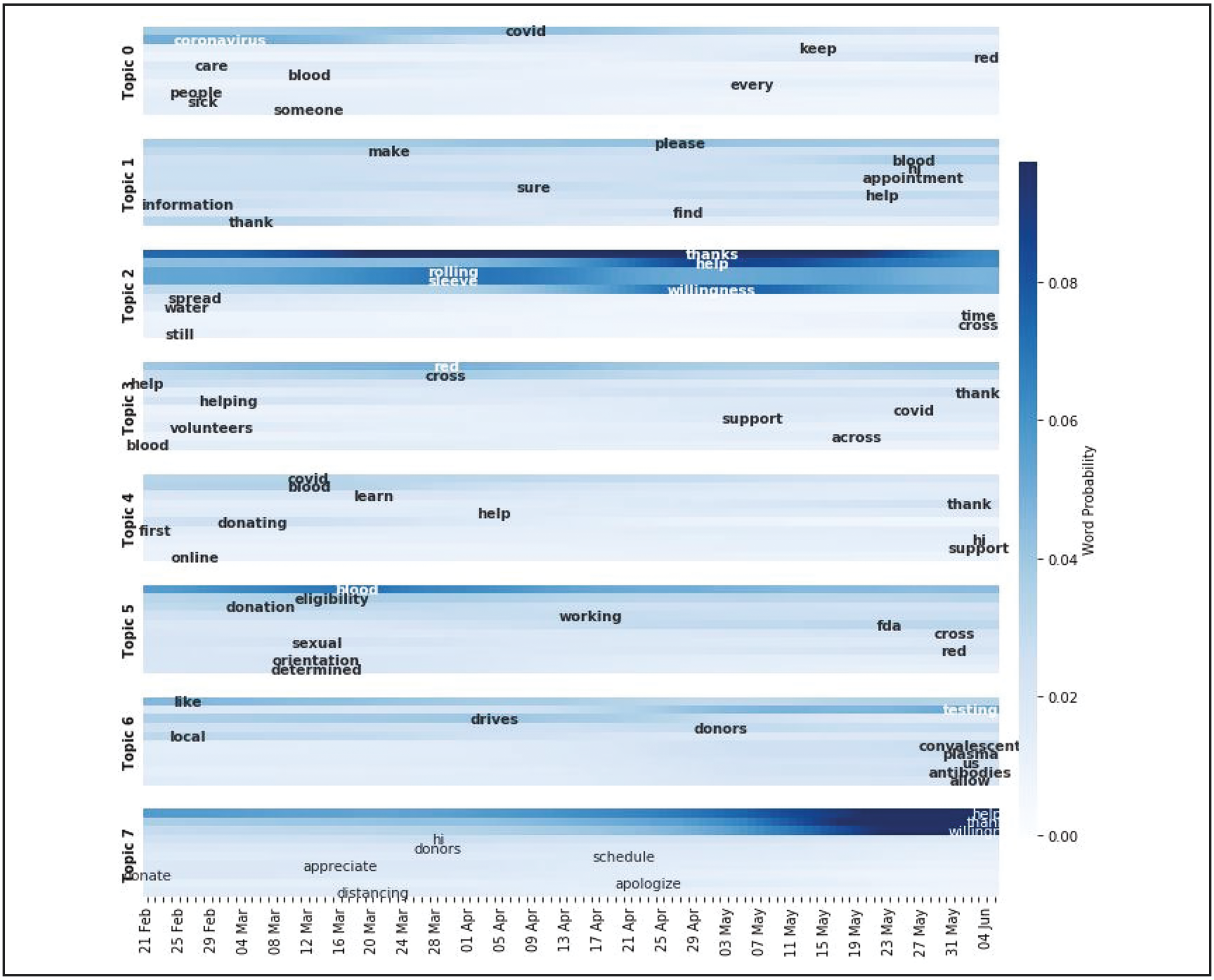

Figure 5. Word probability over time in topics from American Red Cross tweets.

inflammatory bowel disease (IBD) antibodies consistently during the timeline, workers' health and patient care from April, symptoms and hygiene aspects of COVID-19 such as hand washing with soap for at least 20 seconds in March, and the importance of social distancing in March and April. The most unique topic discussed by Mount Sinai is the role of healthcare podcast to share health information. However, some of the hashtags such as #mountsinaistrong and #fitforthefrontlines also became popular.

By interpreting Figures A11 and A12 together, the negative sentiments are observed when the surge of cancer patients is reported, and the lack of social distancing as well as patient care was emphasized. Discussion on need of hygiene practice also resulted in negative sentiment in twitter. Then, positive sentiments are found while importance of frontline support, workers' health, IBD antibodies, and the role of healthcare podcast to share health information became more prominent. In Figure 7, the word probabilities over the time of each topic are replicated.

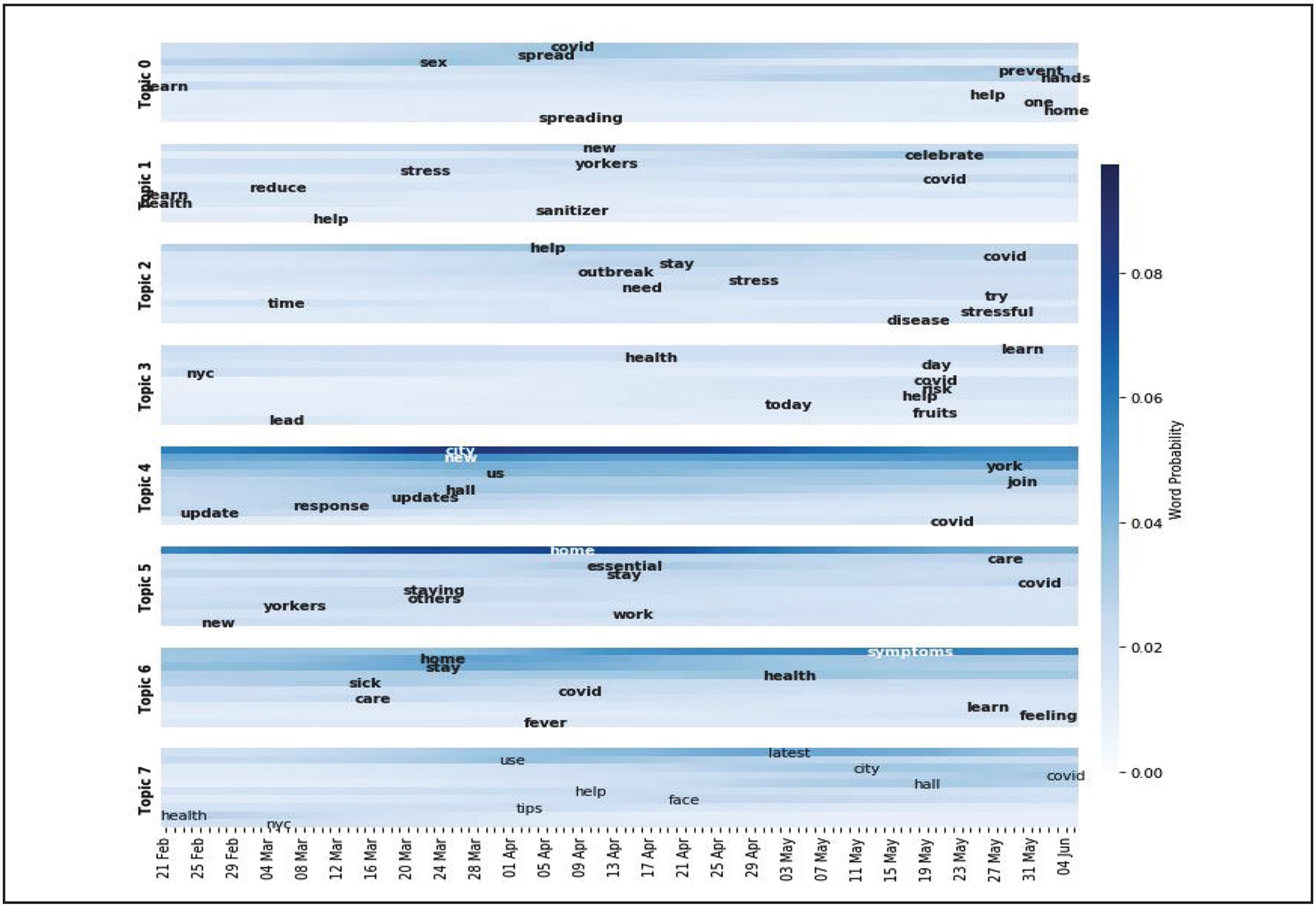

Figure 6. Word probability over time in topics from NYC Department of Health tweets.

COMPARATIVE ANALYSES

In order to identify the specific interaction patters along with the sentiments that might influence the changes (increase and decrease) in the number of cases and deaths in the United States, the topic dynamics and sentiments over time are represented along with the number of cases and deaths of COVID-19 from Worldometer data[31] in Figures 8 and 9.

According to these figures, it can be observed that the document's topic frequency is strengthened at the peak of COVID-19 cases. Furthermore, "willingness to help" is at its highest value after approximately 1 month after the peak of the cases. This may show societal desire to help each other after understanding the consequences of the COVID-19 pandemic. Another interesting observation is the peak in COVID-19 testing's document's topic frequency and the peaks in the numbers of reported cases.

From both Figures 8 and 9, it is obvious that the number of cases and deaths started to increase rapidly from March 20, 2020 and hit the first pick (around 40,000 cases and 2,500 deaths per day) around April 10, 2020. In this time duration, WHO discussed more about the lack of social measures, responses, requirement of adequate number of nurses, and importance of wearing face masks; CDC emphasized on need of contact tracing, failure of preventing the spread, and suggested to follow their guidelines; FEMA mentioned about shortage of medical supplies, responses, and the importance of critical information dissemination.

However, American Red Cross stated about the care of sick people due to COVID-19, volunteer

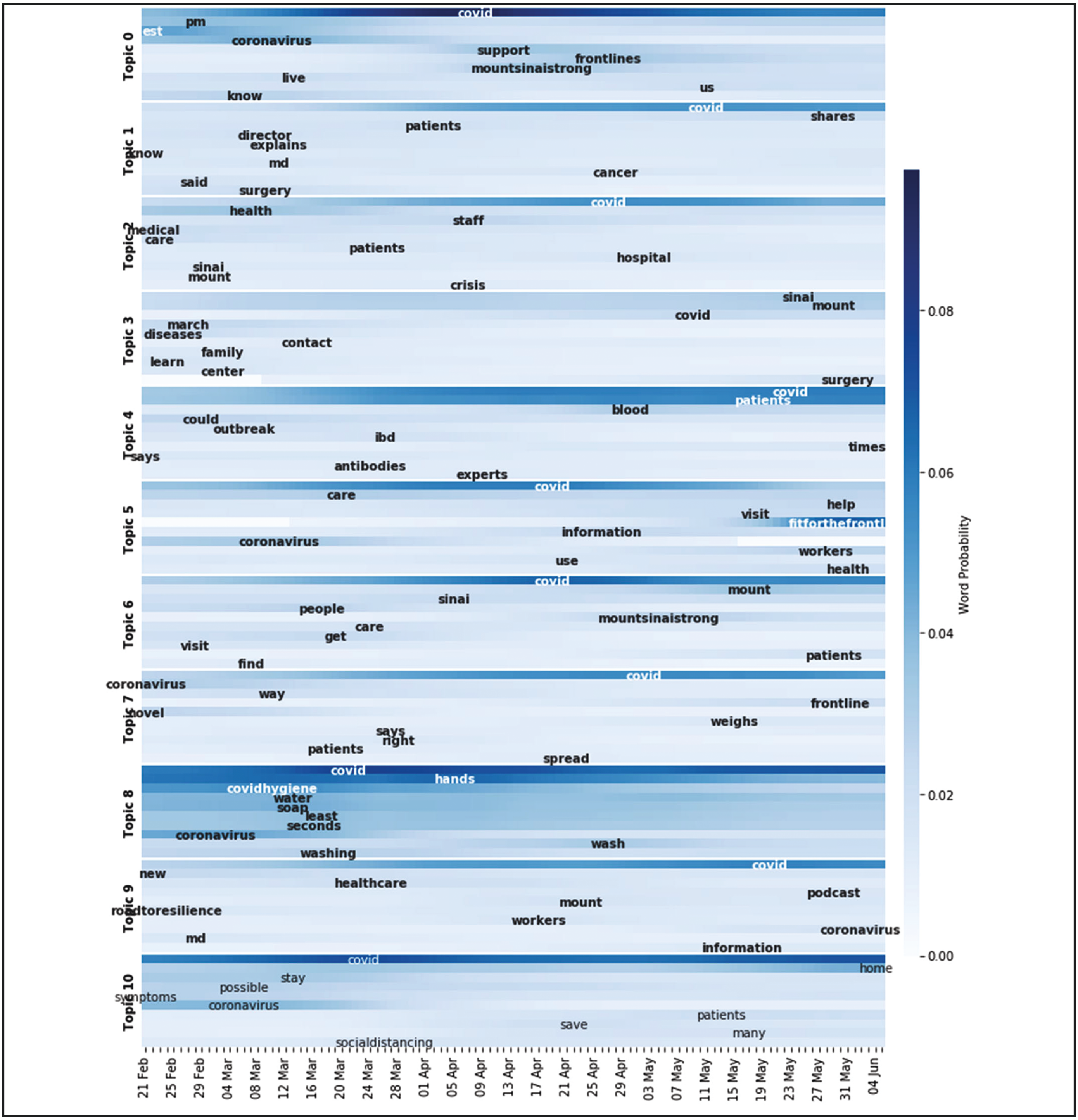

Figure 7. Word probability over time in topics from Mount Sinai tweets.

support, blood donation, and the requirement from FDA to donate blood; NYC Department of Health emphasized on health risks, responses, providing health tips, symptoms of COVID-19, importance of home quarantine, and Mount Sinai Health System focused on patient care, hygiene activities, contact tracing, and healthcare podcast during the mentioned timeline. These figures also represent the similar

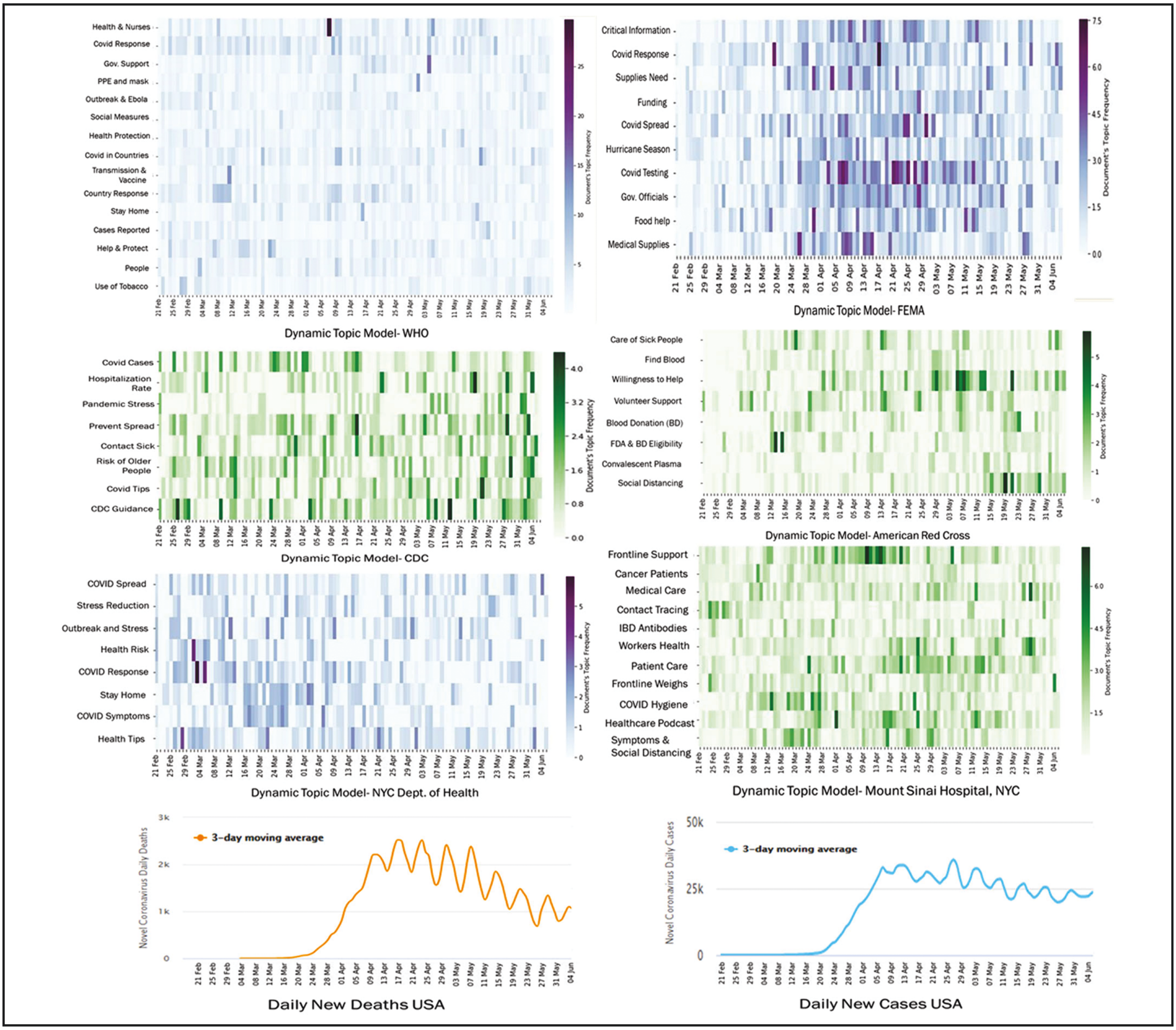

Figure 8. Organizations' topic dynamics with daily new cases and deaths in the United States.

fluctuating patterns in all these organizations over time as well as the number of COVID-19 new cases and deaths. We conclude that society's sentiment in response to the verified news has a short-term component, which is highly volatile and changes with daily news and updates about an infectious disease, in this case, COVID-19-related news.

Then, the number of cases and deaths remained nearly constant from April 11, 2020 to May 10, 2020. In this time, importance of social distancing, increase in funding, response and support from government, essence of pandemic stress management, new measures to prevent the spread, organizations' willingness to help more, frontline support, and the requirement of PPE have started becoming into the focus of the discussion. The probable influence of emphasizing on these issues is obvious from the last portion (May 11, 2020 to June 6, 2020) of the graphs

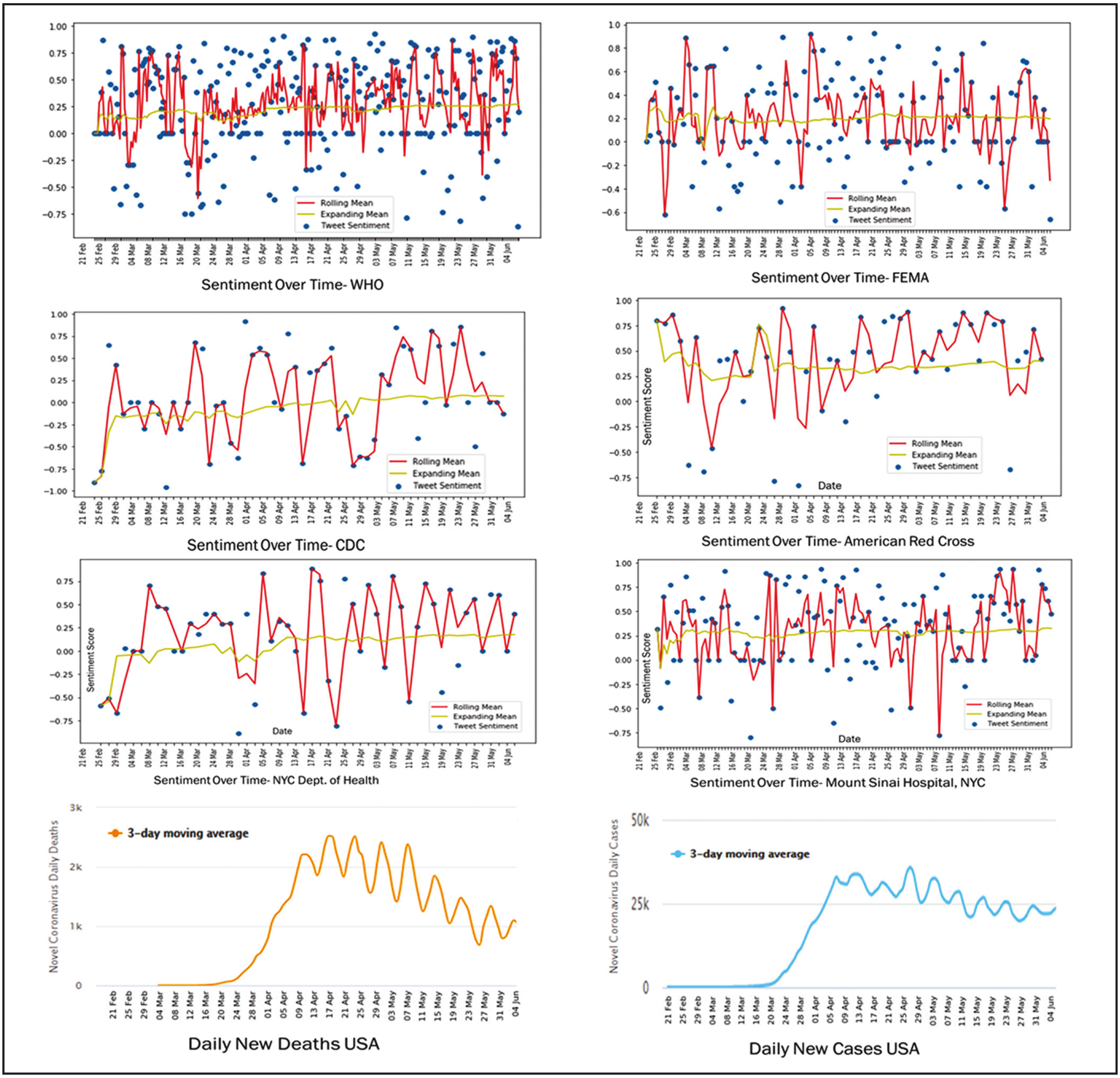

Figure 9. Organizations' sentiment with daily new cases and deaths in the United States.

(Figures 8 and 9), where both the number of cases and deaths decreased. In this timeline, increased response from people and government, awareness about the virus spreading, lowering hospitalization rates, following health guidance, understanding the risk of elderly people, increased medical supply, importance of convalescent plasma donation, medical care, social distancing, working together, pandemic stress management programs, and focusing on workers' health issues also might have influenced human behavior to reduce the number of cases and deaths resulting in flattening the curve.

DISCUSSION OF FINDINGS

The text-based infographics showed in previous sections have revealed different interaction patterns along with the sentiments from the four major public agencies during the pandemic. These interactions were generated from most frequently appeared words in different topics from these entities which are listed later in different timeline. The entire time is divided into four subperiod based on the lockdown (mid-March 2020) and reopening (beginning of the May 2020) time in the United States.

After analyzing all the results and infographics, the noteworthy findings from each organization are listed as follow:

- The unique topics WHO discussed are the importance of PPE, wearing face masks, home quarantine, need of nurses and vaccine, emergence of community transmission, and both effect and use of tobacco.
- The social distancing phrase changed to physical distancing over the time, which indicates the increased awareness and better understanding about pandemic.
- CDC specifically mentioned about the importance of contact tracing, pandemic stress management, and the negative impact of increased hospitalization rates.
- FEMA is the only organization who emphasized on the importance of COVID-19 testing, lack of medical supplies, and the emergence of upcoming hurricane season along with the pandemic.
- American Red Cross put importance on social distancing, blood donation, the eligibility requirement of FDA, and the treatment of COVID-19-infected patients with convalescent plasma.
- Ney York Department of Health and Mental Hygiene emphasized on the pandemic stress management and provided health tips to fight COVID-19.
- Mount Sinai Health System uniquely discussed about the frontline support, IBD antibodies, and the role of healthcare podcast for health information sharing.
- All the organizations discussed about the increased number of COVID-19 cases, need of responses from both people and government, and spreading of the virus.

To identify the interaction patterns that might influence the changes in the number of daily cases and deaths due to COVID-19 are listed as follows as well as in Table 1:

- Before the lockdown time (February 21, 2020 to March 19, 2020), organizations discussed about community transmission, spreading of the virus, contact tracing, emergency management, care for sick people, the need for blood donation along with its FDA requirement, hygiene maintenance, health concerns, and risks along with the need of doctors.
- Just after the lockdown (March 20, 2020 to April 10, 2020), organizations emphasized on the shortage of nurses, medical supplies, social distancing, contact tracing, frontline support, need for volunteer support, and healthcare podcast that might have influenced to the sudden jump in cases and deaths.
- During the lockdown (April 11, 2020 to May 10, 2020), public agencies highlighted about the importance of social distancing, pandemic stress management, PPE requirement, patient care, organizations' willingness to help, stress induced by the outbreak, and the need of government support to manage the pandemic.

Table 1. Most common words from organizations in specific timelines				
Organizations	**February 21, 2020 to March 19, 2020**	**March 20, 2020 to April 10, 2020**	**April 11, 2020 to May 10, 2020**	**May 11, 2020 to June 6, 2020**
	Before the lockdown	Beginning of the lockdown	During the lockdown	Reopening phase
WHO	Global, media, support, minister, personal, covid, increase, protect, health, safe, measures, countries, solidarity, community, transmission, cases, stop, now, spread, safehands, response, prevent	Coronavirus, protective, medical, equipment, nurses, midwives, violence, economic, social, support, stay, home, confirmed, masks	World, international, response, fight, prime, emergency, advice, committee, tools, Europe, physical, active, information, virus, infection	Health, outbreak, Ebola, human, public, tobacco, use, Healthyathome, vaccines, solidarity, pandemic
CDC	Health, US cases, doctor, people, risk, hands, coronavirus, CDC, help	Facebook, coca, spread, older, travel, home, learn, CDC	COVID, age, protect, home, help, gas, spread, states, report	Cases, hospitalization rates, COVIDview, pandemic, stress, slow, guidance, learn, help, protect
FEMA	Information, critical, need, emergency, firefighters, support, disaster, emotional, risk, emergency management, prevent, state, local, officials	Supplies, need, equipment, social, spread, medical, help, federal, guard, hospital, states, national	People, working, right, team, care, response, soldiers, response, medical, supplies, covid, million, funding	Local, state, million, grants, facts, distancing, hurricane, season, tropical, storm, testing, health, food, national, members
American Red Cross	Blood, donation, information, sick, people, care, spread, water, volunteers, donating, eligibility, distancing, sexual, orientation, covid, coronavirus	COVID, make, sure, red, cross, help, learn, drives, donors, working, rolling, sleeve	Please, find, thanks, help, willingness, support, donors, schedule, apologize	Blood, appointment, help, thank, support, time, across, COVID, FDA, testing, convalescent, plasma, antibodies
NYC Department of Health and Mental Hygiene	Learn, reduce, time, health, help, lead, update, response, health, New, Yorkers, sick, care	Stress, covid, spreading, sanitizer, help, staying, home, fever, hall, tips	Outbreak, essential, work, stay, stress, face, need, health, latest, help	Prevent, hands, help, home, celebrate, stressful, disease, covid, risk, fruits, care, symptoms, feeling, city, hall
Mount Sinai Health System	Test, coronavirus, surgery, medical, care, health, diseases, family, contact, center, outbreak, COVIDhygiene, water, soap, roadtoresilience, symptoms, stay, possible	Patients, ibd, antibodies, staff, crisis, experts, care, hands, washing, healthcare, social distancing	Mountsianistrong, support, frontlines, cancer, hospital, covid, blood, information, spread, wash, workers, save	Shares, surgery, COVID, patients, workers, health, weighs, frontline, podcast, information, home, fitforthefrontlines

- In the reopening phase (May 11, 2020 to June 6, 2020), organizations focused on managing hospitalization rate, risk of older people, workers health, adequate medical supplies, maintaining social distancing, and following the health guidance, which might have influenced to reduce the number of cases and deaths.

- Only WHO discussed about the importance of wearing face mask on April, and FEMA emphasized on COVID-19 testing on May, which should be focused earlier and might have influenced the initial increase of the number of cases and deaths in the United States.

This study identified specific topics in Twitter that could have a probable influence on the change in the number of cases and deaths across the United States. Findings could help policy makers to take better preparation for the next phases of the ongoing pandemic as well as in future pandemic situations.

CONCLUSIONS AND FUTURE RECOMMENDATIONS

Social media is considered as an effective information dissemination platform in recent past. The influence of social media on human behavior and decision making is remarkable. Although the spread of misinformation and information overload challenges the effectiveness of the information propagation through social media,[32] it plays a vital role in crisis communication. Hence, exploring the major public agencies interactions over Twitter and interpreting the results along with crowdsourced data were the main focus of this study. From the collected tweets (around 11,000 tweets) of the six major public organizations, it was obvious that WHO remained most active in social media during the first three and half month of COVID-19 pandemic. Topic frequency distribution and sentiment analysis over time helped us to identify the positive and negative influence of different interaction patterns over the time.

The results showed that the importance of wearing face mask was ignored by most of the organizations, although WHO discussed about it in April, which may have been too late. The change of *social distancing* terminology to *physical distancing* over the time indicated the increase in understanding and awareness about COVID-19. However, shortage of nurses, medical supplies, contact tracing, and social distancing were the probable influential factors behind the sudden increase in both the number of cases and deaths in the United States. The importance of COVID-19 testing was only emphasized by FEMA along with the emergence of upcoming hurricane season. CDC focused on pandemic stress management and monitoring the hospitalization rates. American Red Cross put importance on blood donation, the eligibility requirement of FDA, and the treatment of COVID-19-infected patients with convalescent plasma. New York Department of Health and Mental Hygiene emphasized on the pandemic stress management and provided health tips to fight against COVID-19. Mount Sinai Health System focused about the frontline support, IBD antibodies, and the role of healthcare podcast for health information sharing.

The major contribution of this study is the demonstration of prominent discussions from different agencies through social media communication to a larger audience in a major crisis and how they may correlate with community response and societal outcome indicators, ie, growth in number of cases, lockdown activities, among others. Although such outcomes and response depend on several factors, the findings of this research could help the policy makers, government officials, and emergency management organizations to better prepare for the next phases of the ongoing COVID-19 as well as future pandemic. For example, this study identifies specific topics that could potentially influenced the increase and decrease in the number of cases around the globe, ie, information contagion. Decision makers as well as the people in the vulnerable areas could benefit from such information and data-driven approach to revisit social media communication practices to make the strenuous pandemic situation better.

One limitation of this study is in the dataset used being limited to 3,200 historical tweets per agencies based on Twitter Search API limitations. Future studies should utilize more advanced API to conduct more rigorous analysis of how users who follow these agencies reacted to a given information. However, all the machine learning and NLP models (DTM and sentiment analysis) used in this study are probabilistic model; future studies may consider interpretable machine learning models, ie, decision trees,[33] neural network,[34] and random forest,[35] to predict the future trend of discussions in social media. Another emerging issue in social media is the spread of misinformation,[36] which itself became a pandemic, where defining the misinformation[37] should be the first and foremost task. The spread of misinformation also depends on the social network characteristics (followee–follower and user mention),[14,38,39] which needs to be considered to track the misinformation propagation. However, the identification of the good and bad actors in social media can also contribute to tackle the misinformation problem.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-2027360. However, the authors are solely responsible for the findings presented in this study. The analysis and results section are based on the limited Twitter dataset and authors' opinion and cannot be expanded to other datasets without detailed implementation of the proposed methods. Other agencies and entities should explore these findings based on their application/objectives before using these findings for any decision-making purpose. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Author contributions: *The authors confirm the contributions to the paper as follows: study conception and design: MAA and AMS; data collection: MAA and AMS; analysis and interpretation of results: MAA and AMS; draft manuscript preparation: MAA, AMS, and MHA. All authors reviewed the results and approved the final version of the manuscript.*

Md Ashraf Ahmed, PhD Candidate, Department of Civil and Environmental Engineering, College of Engineering and Computing, Florida International University, Miami, Florida.

Arif Mohaimin Sadri, PhD, Assistant Professor, Moss Department of Construction Management, Moss School of Construction, Infrastructure and Sustainability, College of Engineering and Computing, Florida International University, Miami, Florida.

M. Hadi Amini, PhD, DEng, Assistant Professor, Knight Foundation School of Computing and Information Sciences, College of Engineering and Computing, Florida International University, Miami, Florida.

REFERENCES

1. Terhakopian A, Benedek DM: Hospital disaster preparedness: Mental and behavioral health interventions for infectious disease outbreaks and bioterrorism incidents. *Am J Disaster Med.* 2007; 2(1): 43-50.
2. Mortula MM, Ahmed MA, Sadri AM, et al.: Improving resiliency of water supply system in arid regions: Integrating centrality and hydraulic vulnerability. *J Manage Eng.* 2020; 36: 05020011.
3. Ahmed MA, Sadri AM, Hadi M: The role of social networks and day-to-day sharing activity on hurricane evacuation decision consistency and shared evacuation capacity. In *Proceedings of the 99th Annual Meeting of Transportation Research Board*, Washington, DC, USA, 2020.
4. NRC: National Research Council report, disaster resilience: A national imperative, 2012.
5. Chong KC, Cheng W, Zhao S, et al.: Monitoring disease transmissibility of 2019 novel coronavirus disease in Zhejiang, China. *medRxiv*. 2020.
6. Zhang Y, Jiang B, Yuan J, et al.: The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study. *medRxiv*. 2020.
7. Wilder-Smith A, Freedman D: Isolation, quarantine, social distancing and community containment: Pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *J Travel Med*. 2020; 27.
8. Dalton C, Corbett S, Katelaris A: Pre-emptive low cost social distancing and enhanced hygiene implemented before local COVID-19 transmission could decrease the number and severity of cases. Available at SSRN 3549276. 2020.
9. Ahmed MA, Sadri AM, Pradhananga P, et al.: Social media communication patterns of construction industry in major disasters. In *Construction Research Congress Proceedings-2020*. 2020.
10. Roy KC, Ahmed MA, Hasan S, et al.: Dynamics of crisis communications in social media: Spatio-temporal and text-based comparative analyses of twitter data from Hurricanes Irma and Michael. In *Proceedings of the International Conference on Information Systems for Crisis Response and Management (ISCRAM) 2020*. 2020.
11. Mojumder MN, Ahmed MA, Sadri AM: Identifying ridesharing risk, response, and challenges in the emergence of novel coronavirus using interactions in uber drivers forum. *Front Built Environ*. 2021; 7.
12. Brown J: Is social media the key to effective communication during campus emergencies. Govtech. 2015. Available at *http://www.govtech.com/education/Is-Social-Media-the-Key-to-Effective-Communication-During-Campus-Emergencies.html*.
13. Sadri AM, Ukkusuri SV, Ahmed MA: Review of social influence in crisis communications and evacuation decision-making. *Transp Res Interdiscip Perspect.* 2021; 9: 100325.
14. Sadri AM, Hasan S, Ukkusuri SV, et al.: Understanding information spreading in social media during Hurricane Sandy: User activity and network properties. arXiv preprint arXiv:170603019. 2017.
15. Morshed SA, Arafat M, Ahmed MA, et al. Discovering the commuters' assessments on disaster resilience of transportation infrastructure. In *Proceedings of ASCE International Conference of Transportation and Development (ICTD), 2020*. 2020.
16. Zhang C, Fan C, Yao W, et al.: Social media for intelligent public information and warning in disasters: An interdisciplinary review. *Int J Inform Manag.* 2019; 49: 190-207.
17. Austin L, Fisher Liu B, Jin Y: How audiences seek out crisis information: Exploring the social-mediated crisis communication model. *J Appl Commun Res.* 2012; 40(2): 188-207.
18. Freimuth VS, Hilyard KM, Barge JK, et al.: Action, not talk: A simulation of risk communication during the first hours of a pandemic. *Health Promotion Practice*. 2008; 9(4_suppl): 35S-44S.
19. Palen L, Hughes AL: Social media in disaster communication. In Rodríguez H, Donner W, Trainor JE (eds.): *Handbook of Disaster Research*. Berlin: Springer International Publishing, 2018: 497-518.
20. Monahan B, Ettinger M: News media and disasters: Navigating old challenges and new opportunities in the digital age. In Rodríguez H, Donner W, Trainor JE (eds.): *Handbook of Disaster Research*. Berlin: Springer International Publishing, 2018: 479-495.
21. Battur R, Yaligar N: Twitter bot detection using machine learning algorithms. *Int J Sci Res*. 2020; 8: 304-307.
22. Asr FT, Taboada M: The data challenge in misinformation detection: Source reputation vs. content veracity. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Association for Computational Linguistics, 2018: 10-15.
23. Huang S-K, Lindell MK, Prater CS: Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. *Environ Behav.* 2016; 48(8): 991-1029.

24. Agarwal A, Xie B, Vovsha I, et al.: Sentiment analysis of twitter data. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*. 2011: 30-38.
25. Wang X, Wei F, Liu X, et al.: Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM. 2011: 1031-1040.
26. Hutto C, Gilbert E: Vader: A parsimonious rule-based model for sentiment analysis of social media text. 2014.
27. Beri A: Sentimental analysis using VADER—Interpretation and classification of emotions. Available at *https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664#:~:text=VADER%20(%20Valence%20Aware%20Dictionary%20for,intensity%20(strength)%20of%20emotion.&text=The%20sentiment%20score%20of%20a,each%20word%20in%20the%20text*.
28. Blei D: Probabilistic topic models. *Commun ACM*. 2012; 55(4): 77-84. DOI: 10.1145/2133806.213382.
29. Blei DM, Lafferty JD: Dynamic topic models. In *ICML'06: Proceedings of the 23rd International Conference on Machine Learning*. 2006; 113-120.
30. Dyer H: The story of worldometer, the quick project that became one of the most popular sites on the internet. Available at *https://www.newstatesman.com/science-tech/coronavirus/2020/05/story-worldometer-quick-project-became-one-most-popular-sites*.
31. Worldometer: COVID-19 coronavirus pandemic data. Available at *https://www.worldometers.info/coronavirus/country/us/*.
32. Barua Z, Barua S, Aktar S, et al.: Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. *Prog Disaster Sci*. 2020; 8: 100119.
33. Safavian SR, Landgrebe D: A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern*. 1991; 21(3): 660-674.
34. Anthony M, Bartlett PL: *Neural Network Learning: Theoretical Foundations*. Cambridge: Cambridge University Press, 2009.
35. Breiman L: Random forests. *Mach Learn*. 2001; 45(1): 5-32.
36. Chou W-YS, Oh A, Klein WM: Addressing health-related misinformation on social media. *JAMA*. 2018; 320(23): 2417-2418.
37. Wu L, Morstatter F, Carley KM, et al.: Misinformation in social media: Definition, manipulation, and detection. *SIGKDD Explor Newsl*. 2019; 21(2): 80-90.
38. Sadri AM, Hasan S, Ukkusuri SV: Joint inference of user community and interest patterns in social interaction networks. *Soc Netw Anal Min*. 2019; 9(1): 11.
39. Sadri AM, Hasan S, Ukkusuri SV, et al.: Exploring network properties of social media interactions and activities during hurricane sandy. *Transp Res Interdiscip Perspect*. 2020;6:100-143.

APPENDIX A

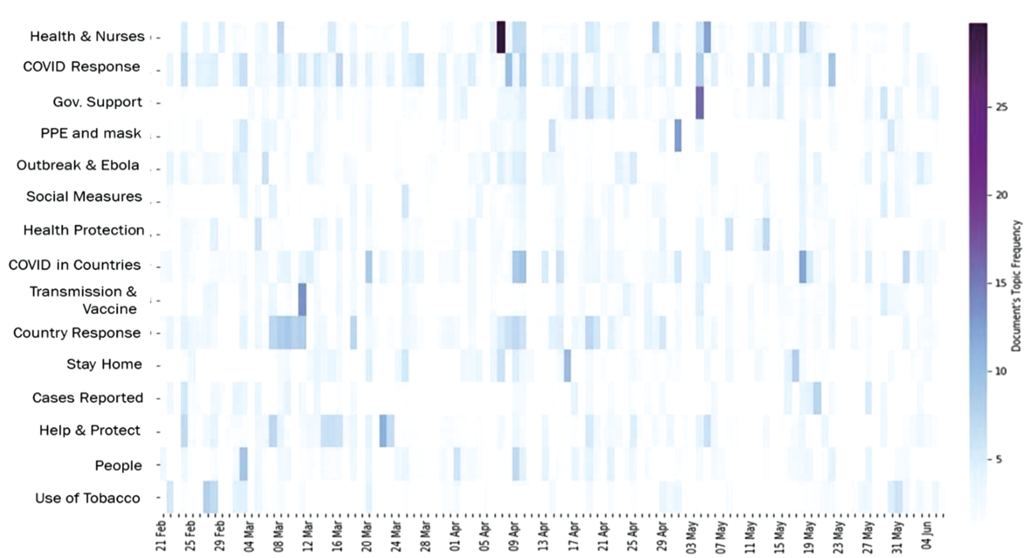

Figure A1. Topic frequency over time from WHO tweets.

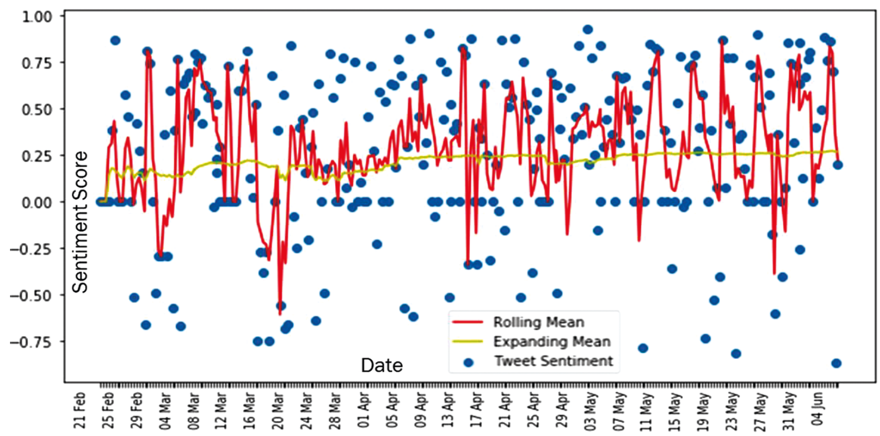

Figure A2. Sentiment analysis over time from WHO tweets.

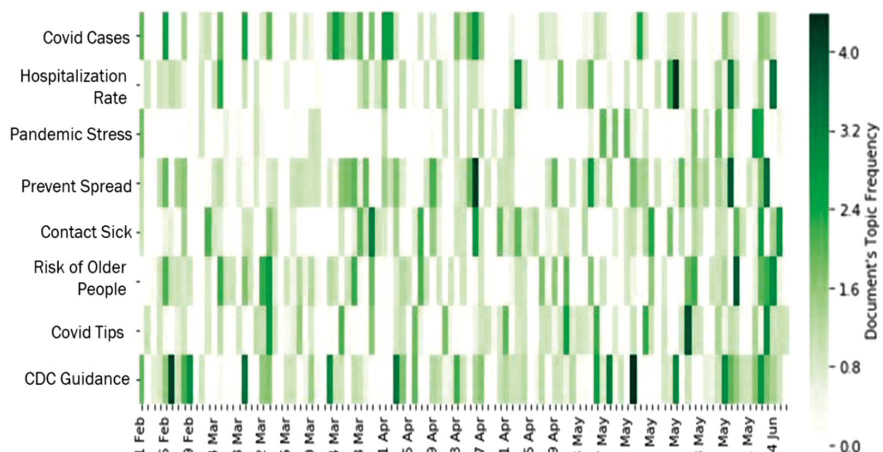

Figure A3. Topic frequency over time from CDC tweets.

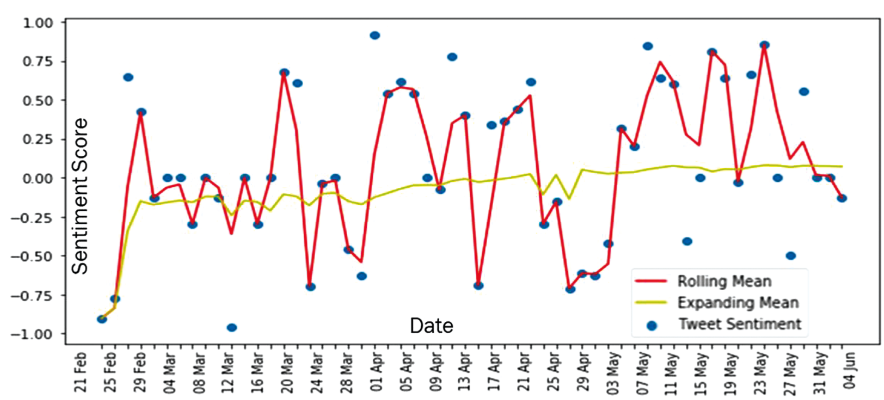

Figure A4. Sentiment analysis over time from CDC tweets.

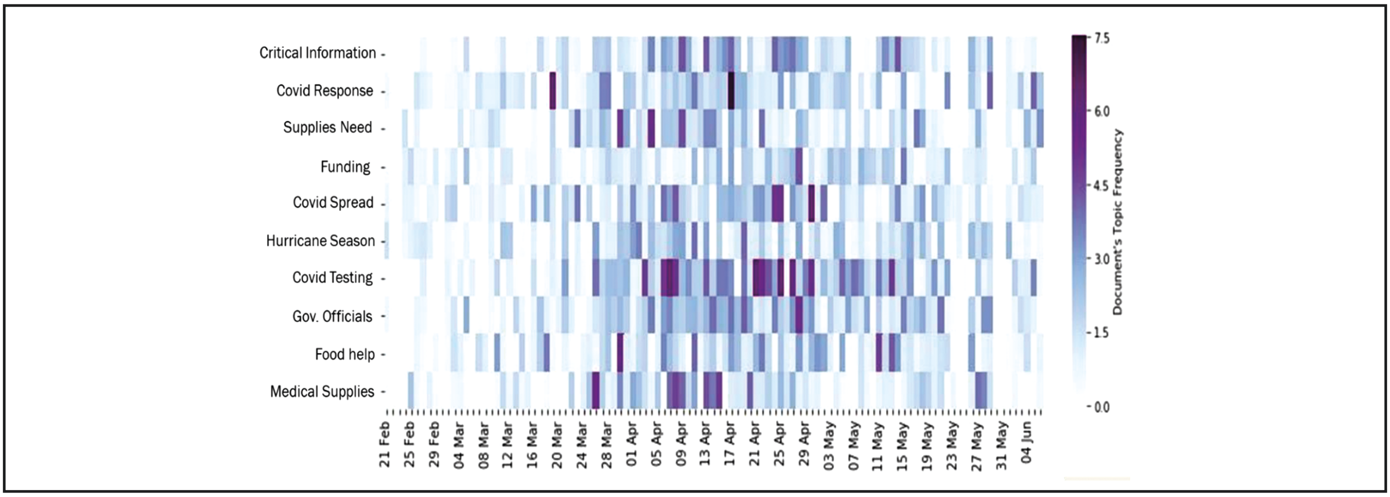

Figure A5. Topic frequency over time from FEMA tweets.

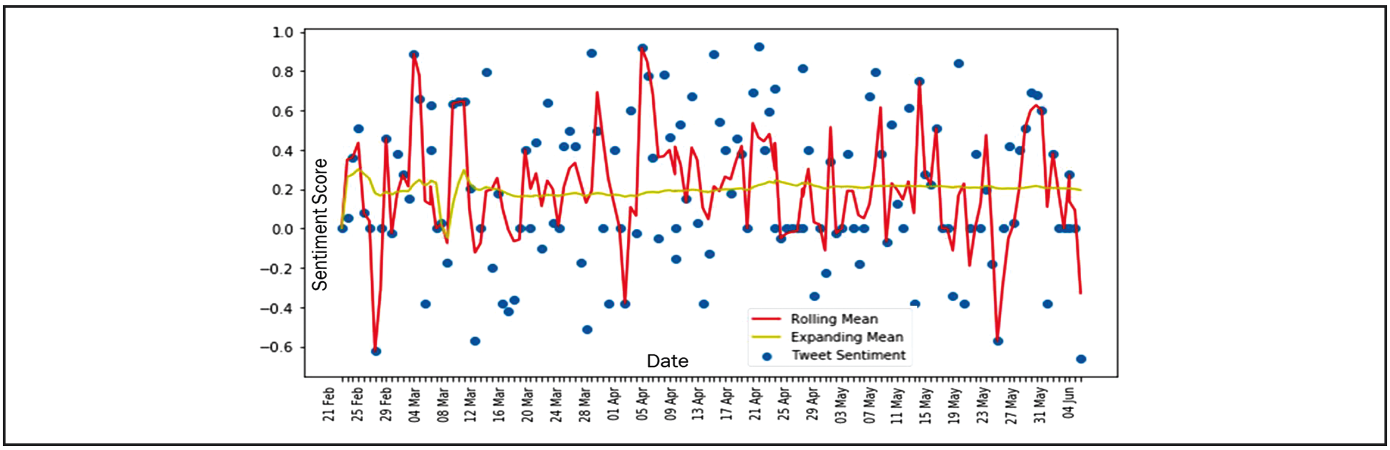

Figure A6. Sentiment Analysis over time from FEMA tweets.

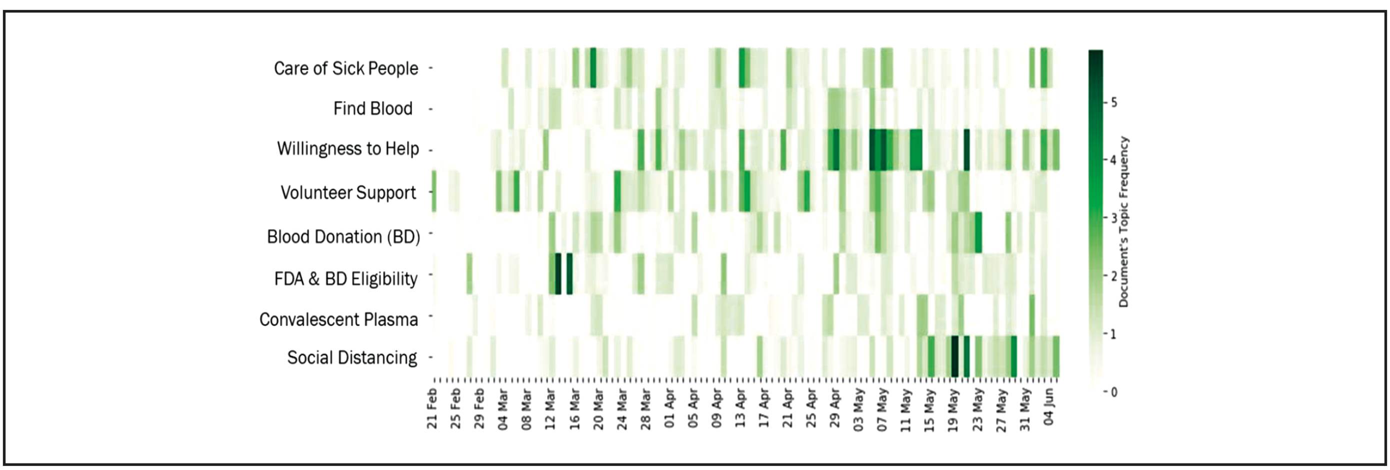

Figure A7. Topic frequency over time from American Red Cross tweets.

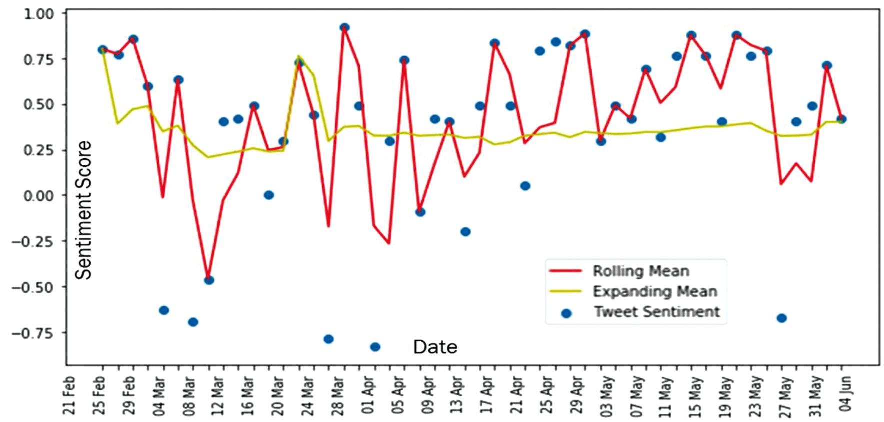

Figure A8. Sentiment analysis over time from American Red Cross tweets.

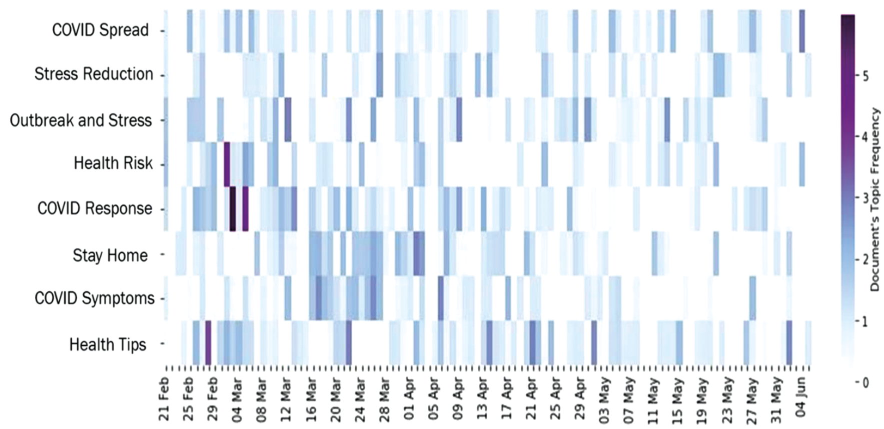

Figure A9. Topic frequency over time from NYC Department of Health tweets.

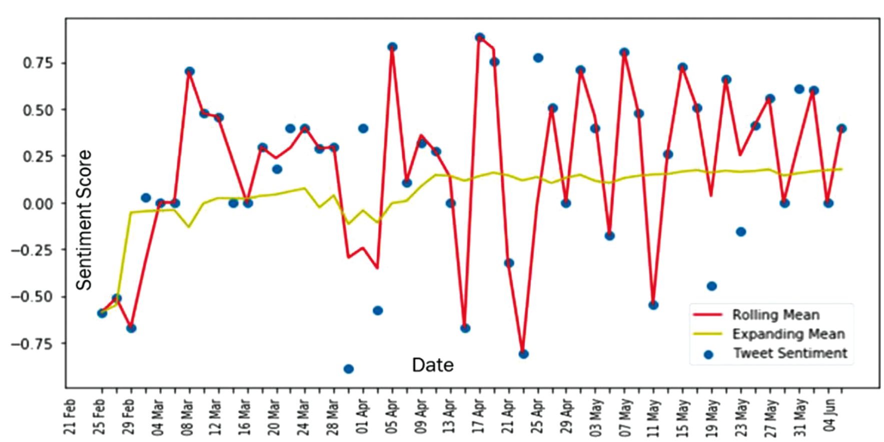

Figure A10. Sentiment analysis over time from NYC Department of Health tweets.

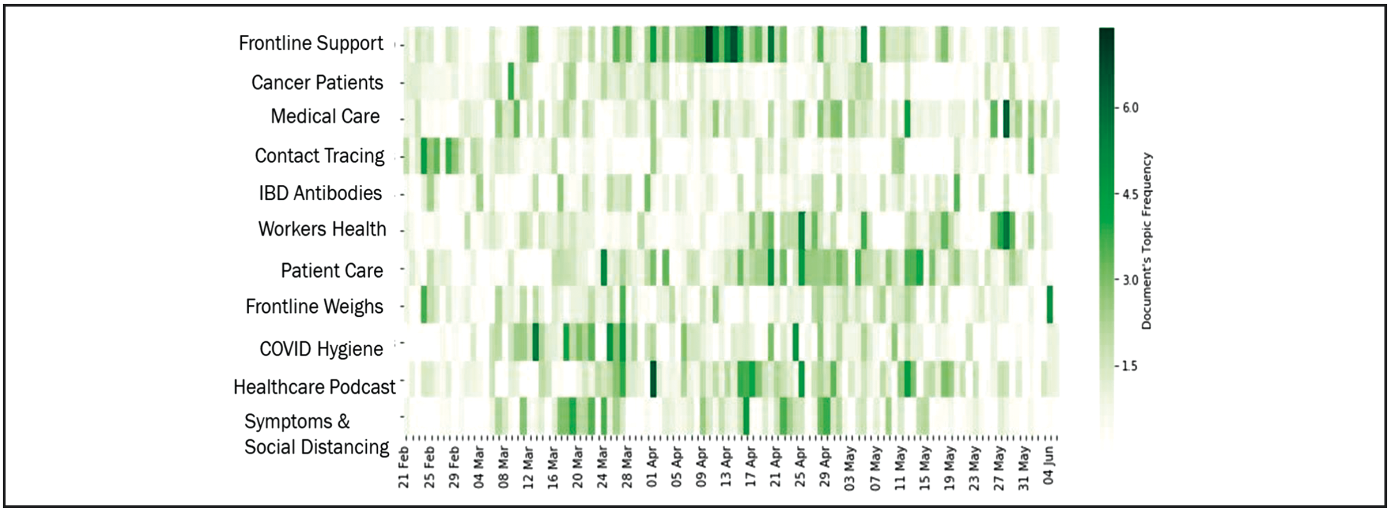

Figure A11. Topic frequency over time from Mount Sinai tweets.

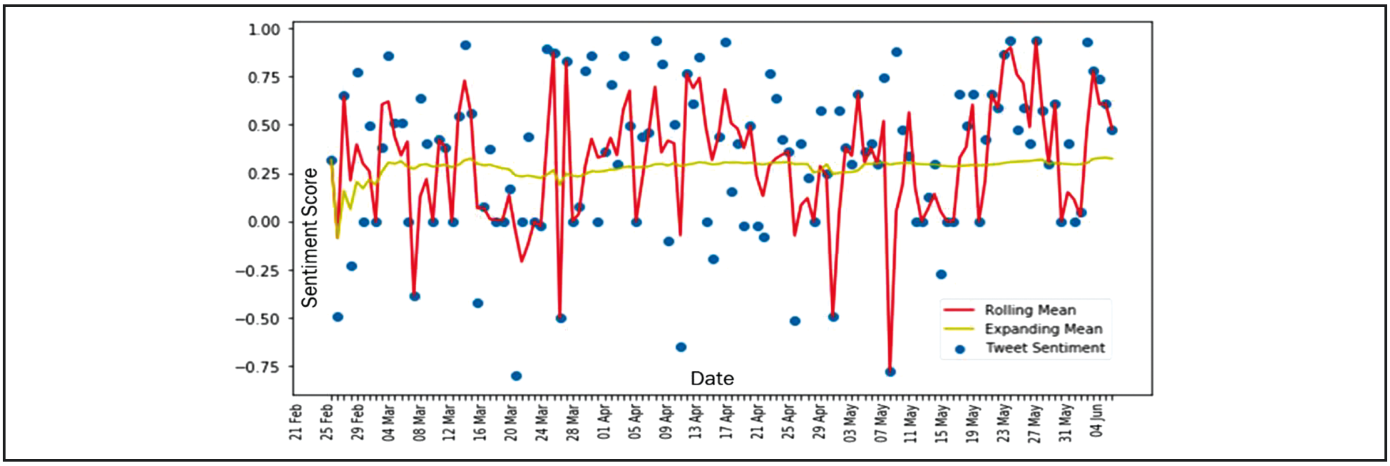

Figure A12. Sentiment analysis over time from Mount Sinai tweets.