MEDIA & COMMUNICATION STUDIES | RESEARCH ARTICLE

What public health campaigns can learn from people’s Twitter reactions on mask-wearing and COVID-19 Vaccines: a topic modeling approach

Yi (Jasmine) Wang1*, Molu Shi2 and Jueman (Mandy) Zhang3

Abstract: Topic modeling, which uses machine learning algorithms to identify the emergence of topics, can help public health professionals monitor online public responses during health crises. This study used Latent Dirichlet Allocation algorithm to model the topics in Twitter messages (or “tweets”) from the US during the COVID-19 pandemic from March 20th to August 9th, 2020. Topic sizes and sentiment were calculated as the pandemic evolved, for major topics about vaccination and mask-wearing as a nonpharmaceutical intervention measure. Despite the pandemic, positive sentiments were found among most topics. While users were found to react more often to positive sentiment about mask-wearing, negative content on vaccination was found more popular. Noticeable trends in topic sizes and sentiment were observed for various topics, which correlated in time with some key pandemic events and policy changes, implying their impacts on social media responses. By analyzing such trends and impacts, this research offers insights on health campaign message design and how to outreach the general public most effectively.

Subjects: Risk Communication; Communication Research Methods; Health Communication; New Media

ABOUT THE AUTHOR

Yi (Jasmine) Wang (Ph.D., University of Connecticut) is an assistant professor of communication at the University of Louisville. Her research interests focus on information processing and design, as well as the role of social media in communicating health and risk information. Molu Shi (Ph.D., Massachusetts Institute of Technology) is a data scientist in healthcare based in Louisville, KY. His research interests focus on natural language processing in healthcare. Jueman (Mandy) Zhang (Ph.D., Syracuse University) is an associate professor in the Department of Communication Arts at New York Institute of Technology. Her research areas include health and crisis communication, digital media, and journalism. This project was initiated and carried along by our shared interest in risk communication in healthcare.

PUBLIC INTEREST STATEMENT

COVID-19, the ongoing infectious disease, has been reported to be the worst global pandemic in a century. This study examined the social media reactions to vaccination and wearing masks by analyzing Twitter messages (or tweets) using a topic modeling approach. We identified five most discussed topics about vaccination and mask-wearing, respectively. We also analyzed the trends in topics and sentiment in relation to key events and policy changes. Twitter discussions on mask-wearing and vaccination were generally positive, except for the science related topic on vaccination during the study period (March 20th to August 9th, 2020). People also retweeted more positive content about mask-wearing but more negative content about vaccination. Our study emphasized the important public health agenda setting roles of government and media for its influence on social information transmission on Twitter during the COVID-19 pandemic. The study also provided insights on how to design and time health campaigns more effectively.

© 2021 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.
Keywords: topic modeling; LDA; public health campaign; vaccine; mask-wearing; emotion; sentiment; popularity

COVID-19, an ongoing infectious disease caused by a novel coronavirus, has been reported the worst global pandemic in a century (WHO, n.d.). Although the pandemic is expected to end with a herd immunity established through vaccination, continued adherence to non-pharmaceutical interventions (NPIs) is still needed to slow the spread (Patel et al., 2020), prior to significant vaccination coverage. Among all NPI measures, wearing masks could be implemented with low cost and without radical societal disruptions (Li et al., 2020). This study examined the social media reactions to vaccination and wearing masks by analyzing Twitter messages (or tweets) using a topic modeling approach. In particular, we analyzed the trends in topics and sentiment in relation to key events and policy changes. Our study emphasized the important public health agenda setting roles of government and media for its influence on social information transmission on Twitter during the COVID-19 pandemic. The study also provided insights on how to design and time health campaigns more effectively.

1. Literature review

1.1. Compliance to NPIs and attitude towards vaccination

Before sufficient vaccination coverage, continued compliance to NPIs is needed to reduce the spread (Patel et al., 2020). Measures of NPIs have reportedly been temporarily effective in reducing COVID-19 transmission in countries such as China (Pan et al., 2020) and historically effective in the 1918–1919 influenza pandemic in the United States (Hatchett et al., 2007). However, the impact of NPIs depends largely on how well they are implemented. While a CDC survey reported 80% of the participants complying with social-distancing (CDC, 2020, June 12), studies showed a lower compliance rate on mask-wearing: 65% and 44% of the participants reported wearing masks all or most of the times in enclosed spaces, on themselves and on other people, respectively (Igielnik, 2020, June 23, June 23).

While COVID-19 vaccines were being developed in 2020, public attitudes posed challenges to advocate vaccination. A Gallup study (O’Keefe, 2020, Aug. 7, August 7) revealed that 35% of Americans are reluctant to get a free FDA-approved vaccine. An Axio-Iposos study (July20th, 2020) showed that 61% of the Americans perceive the first-generation vaccine with moderate or high risk.

To increase population acceptance of vaccination and engagement with NPIs including mask-wearing, public health campaigns have been implemented, especially on social media as a critical campaign channel (HHS, n.d.). Therefore, knowledge of what people have discussed on social media regarding vaccination and mask-wearing could help design more effective public health campaigns.

1.2. Topic modeling for social media discourse during public health crises

As a major social media platform and a pivotal source for text based public discourse, Twitter has been studied to understand public reactions during various public health crises, including the Covid-19 pandemic (e.g. Abd-Alrazaq et al., 2020; Doogan et al., 2020; Kwok et al., 2021; Liu et al., 2020). Research on the role of Twitter in face of public health crises has focused on two directions: (a) Twitter as an information source for emergency responses management and (b) Twitter as the collective outcome for public perceptions and attitudes about health and risk issues. On one hand, public health and government agencies such as the Centers for Disease Control and Prevention (CDC), the National Institutes of Health (NIH), and governments at different levels have been using Twitter to provide updates and key information to the public. Research in this realm has focused on how to use social media to manage information dissemination and improve health and risk communication engagement by a wider audience (e.g. Carley et al., 2016; Xu, 2020). On the other hand, the participatory nature of social media generates proliferation of real-time
geolocated user data, which amplifies the role of the public in creating trends, and provides opportunities for researchers and stakeholders in public health to identify risk factors (e.g. Paul & Dredze, 2011) and to monitor the public perception of health crises (e.g. 2009 H1N1 pandemic in Chew & Eysenbach, 2010; Ebola outbreak in Odlum & Yoon, 2015; HPV vaccine in Keim-Malpass et al., 2017). Following the latter direction, we aimed to identify the collective public perceptions and attitudes about (1) mask-wearing and (2) vaccination that are reflected on Twitter.

Topic modeling is a powerful methodology to structure massive and unstructured textual social media data into topics (e.g. Cambria et al., 2013). While traditional methods such as content analysis can identify topics using human coders, they can be time consuming and of low reliability (Günther & Domahidi, 2017). Topic modeling, on the other hand, uses a machine learning approach to automatically analyze textual data and determine topics using algorithms, such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003). The automotive nature of such method is critical, especially during a fast-evolving pandemic, where quick responses from the public health agencies are expected.

Topic modeling has been used in recent research to analyze media content of the COVID-19 pandemic. While most research identified general topics related to the COVID-19 outbreak such as origin, prevention, and impact (e.g. Abd-Alrazaq et al., 2020; Liu et al., 2020), only a few focused on non-pharmaceutical interventions (NPIs) (Doogan et al., 2020) or vaccination (Kwok et al., 2021). Therefore, we intended to study Twitter discourse on two preventive measures: mask-wearing and vaccination, with a particular focus on the United States, where the highest number of confirmed cases and deaths were confirmed (John Hopkins Coronavirus Resource Center, n.d.), and unsatisfactory compliance rate for mask-wearing (Igielnik, 2020, June 23, June 23) and moderate hesitancy level for vaccination were revealed in earlier surveys (O’Keefe, 2020, Aug. 7, August 7). Topic modeling will help discover the underlying topical patterns regarding the two preventive measures: what and to what extent people have discussed mask-wearing and vaccination on Twitter and how the topics evolved across the study period. The findings will shed light on what topics to be addressed in health campaigns to effectively promote the two preventive measures.

Thus, we ask the following research questions:

Research question 1 (RQ. 1): What were the most discussed topics about (1) mask-wearing and (2) vaccination on Twitter from March 20th to August 9th, 2020, in the United States?

Research question 2 (RQ. 2): How did people’s discussions on topics about (1) mask-wearing and (2) vaccination on Twitter change over time during the study period?

1.3. Sentiment analysis of social media discourse and social transmission

Sentiment analysis has been increasingly used to analyze text about health-related issues (e.g., Bantum & Owen, 2009; Cabling et al., 2018). Recently, it has been used to study Twitter discourse during the early stages of the COVID-19 pandemic (e.g., Bhat et al., 2020; Chandrasekaran et al., 2020). Sentiment analysis is a text classification methodology that automatically determine valence (i.e., positive, neutral or negative) and emotions from text by comparing a message to a previously calibrated sentiment lexicon—a collection of opinion words (e.g. beautiful, amazing) (Liu, 2012). Because of its ability to automate the analyses on massive amounts of data, sentiment analysis offers great opportunities to study the vast, opinionated data from social media to understand people’s attitudes towards and emotions about a topic (Liu, 2012). When used together with topic modeling, sentiment analysis will help us understand how the valence or the prevailing emotion is associated with different topics. To date, limited is known about the overall valence of Twitter discourse about specific NPI measures such as mask wearing. Kwok et al.’s (Kwok et al., 2021) sentiment analysis of Australia-based tweets revealed that two-thirds of the tweets about COVID-19 vaccination were positive and the remaining one-third were negative. However, there is a paucity of knowledge about how such difference in valence can be attributed to the difference in their underlying topics or how it evolved over time. This study
aimed to fill the gaps. We identified the valence of each topic about the two preventive measures: mask-wearing and vaccination, and also investigated how the valence evolved across the study period. The findings will provide guidance on how to affirm people’s positive attitudes and how to address people’s negative concerns. According to the abovementioned literature, we asked:

Research Question 3 (RQ. 3): What was the valence for each of the topics detected about (1) mask-wearing and (2) vaccination on Twitter from March 20th to August 9th, 2020 in the United states?

Research Question 4 (RQ. 4): How did the valence on (1) mask-wearing and (2) vaccination on Twitter change over time during the study period?

Furthermore, social media has been a powerful tool for information transmission, especially during the COVID-19 pandemic, when people are encouraged to stay at home with limited physical social interactions. Because the abundance of information competes for attention and influences public attention, public health campaigns face a barrier due to their limited reach (Snyder & Hamilton, 2002). As a result, it is vital to understand the determinants of social transmission of health-related information to facilitate the virality of health campaigns.

Multiple studies have explored the factors contributing to the virality of social media information (e.g., Akpinar & Berger, 2017; Hansen et al., 2011; Nikolakou & King, 2018; So et al., 2016; Tellis et al., 2019). In general, researchers agree that emotion drives online sharing motivations in most contexts (Berger & Milkman, 2012; Stiegitz & Dang-Xuan, 2013; Tellis et al., 2019). While lay belief on news diffusion suggested that negative news passes along more easily than positive news (Godes et al., 2005), large scale content analyses of different types of media content show that positive content that elicits positive emotions such as excitement and inspiration is more likely to be shared on social media (video news in Al-Rawi, 2019; news articles in Berger & Milkman, 2012; personal tweets in Bollen et al., 2011; commercials in Tellis et al., 2019; health news in Kim, 2015). These studies took a holistic approach rather than focusing on media content of specific issues. In response to the new virus, vaccination under clinical trials could be perceived as of more uncertainty than the NPI measure of mask-wearing. Given such difference, this study intended to compare them in terms of social transmission strength, in relation to their valences. Therefore, we ask the following research questions:

Research Question 5 (RQ. 5): Did people retweet and react more to positive or negative content on Twitter for (1) mask-wearing and (2) vaccination during the study period?

1.4. Agenda setting

One framework to study public reactions on social media is rooted in the theories of agenda setting (McCombs & Shaw, 1972) and intermedia agenda setting (McCombs, 2004). The original agenda setting theory, which is referred to as the first level agenda setting, posits the transfer of salience of issues from news media to the public agenda (McCombs & Shaw, 1972). Its extended version, which is referred to as the second-level agenda setting, postulates the transfer of salience of attributes and their affective tones (Coleman et al., 2008). While the agenda-setting theory is often evidenced in political contexts (e.g. Conway-Silva et al., 2018), agenda setting has been applied to study health issues (e.g. Ogata Jones et al., 2006; Pratt et al., 2002) and other public crises (e.g. climate change in Su & Borah, 2019; earthquake in Valenzuela et al., 2017). Intermedia agenda setting further expands the theory by examining the interactions between different media outlets, for instance, how traditional media and Internet-based media such as social media interact and shape each other’s agenda (McCombs, 2004). Previous research on agenda setting on Twitter suggests that although the influence between traditional media and social media is reciprocal, in general, traditional media still has greater top-down agenda-setting power (e.g. Conway-Silva et al., 2018; Meraz, 2009).
Past research has identified factors that can impact the strength of the agenda setting effects. From an audience’s perspective, for instance, their need for orientation, an audience characteristic driven by issue relevance and uncertainty, was found to enhance the agenda-setting effects (Coleman et al., 2008). That is, when people feel a lack of information about an issue (high uncertainty), which they perceive as being personally or socially important (high relevance), their need for orientation increases, resulting in a stronger agenda setting effect. From the media perspective on the other hand, the level of issue obtrusiveness was found to influence the agenda setting effects (Coleman et al., 2008). For unobtrusive issues, about which people do not have direct and personal experience (Zucker, 1978), agenda setting effects are stronger since people rely on media coverage to understand such issues.

For the general public, as the COVID-19 pandemic has affected all aspects of life, it can be perceived as of high personal and social relevance. In addition, COVID-19 is a new coronavirus, about which knowledge is limited, thus posing high uncertainty. Therefore, the public’s need for orientation would be high. Moreover, the novelty and evolving nature of the pandemic makes the issue unobtrusive. As a result, the agenda setting effects are expected to be amplified during the pandemic. As the COVID-19 pandemic evolves, changes in the infectious rate, public policies, and health recommendations can all impact people’s reactions on social media. This study intended to examine if Twitter discourse on (1) mask-wearing and (2) vaccination were affected by news media coverage of key events, such as the White house’s plan to reopen the economy (Edelman et al., 2020, April 16 April) and the CDC’s suggestion to reopen schools (Goodnough, 2020 July). The findings will further our understanding of the influence of the media and the government in shaping the public agenda and the sentiment regarding the two preventive measures as the pandemic evolved. Therefore, we ask:

Research question 6 (RQ. 6): How did the evolution of the pandemic and change in public policy influence people’s discussions on topics about (1) mask- wearing and (2) vaccination on Twitter during the study time period?

Research question 7 (RQ. 7): How did the evolution of the pandemic and changes in public policy influence the sentiment change on (1) mask- wearing and (2) vaccination on Twitter during the study period?

2. Methods

2.1. Source data
We used 154,978 geo-tagged coronavirus tweets from March 20th to August 9th, 2020 (Lamsal, 2020), first-place winner in the Spring 2020 IEEE DataPort Dataset Competition. This source data contained tweet id’s and sentiment scores calculated via TextBlob (Textblob, 2020), extracted from Twitter using 54 COVID-19 related keywords. While we used the data till August 9th, 2020, daily updates are continuously received, to visualize real-time Twitter feeds on COVID-19 (Live Twitter Sentiment, 2020).

| Table 1. Keywords to subset tweets about mask-wearing and vaccination |
|---------------------------------------------------------------|
| **Corpus** | **Keywords/hashtags** |
| Mask | mask, facemask, fac cover, faceshield weararmask billionshieldschallenge, billionshields, nomaskonne, nomask, namask, maskless, maskup, nanewnormal, nomuzzles, nomaskselfie, showyoursmile, facecoverings |
| Vaccine | vaccine immunity, immune, immunization, pharm |
2.2. Data processing

We re-hydrated the tweet data through their tweet id’s via Twitter API using twarc (Twarc, 2020). To select corpora on US tweets about mask-wearing and vaccination, we subbed the source data using two sets of keywords (see Table 1), resulting in 7,647 and 416 original tweets (non-retweets), respectively. The geo-administration level data from each tweet was reverse obtained through its geocoordinates using reverse_geocoder (Reverse Geocoder, 2020).

2.3. Topic modeling

Latent Dirichlet Allocation (LDA) was used to model topics discussed within the corpora, where topics in a given document and words in a given topic both follow a Dirichlet distribution, characterized by their concentration parameter \( \alpha \) and \( \beta \) (Blei et al., 2003). Given a corpus as training data, the optimal values of \( \alpha \) and \( \beta \) can be fitted for each fixed number of topics \( K \), while maximizing the coherence score \( C_v \) (Röder et al., 2015), where the coherence score \( C_v \) (0 \( \leq \) \( C_v \) \( \leq \) 1) characterizes the coherence, a likelihood measure of word co-occurrence in the same topics. Typically, a higher \( C_v \) implies a better topic model. With a fitted model, each word in a new text sample can be represented with a percentage score associated with each topic, which allows us to generate key statistics to describe data within and beyond the corpus, such as the size of topics, defined by the percentage of words assigned to the topic (see Sec. Measures).

We fitted two LDA models on the corpora about mask-wearing and vaccination, respectively, using gensim (Řehůřek & Sojka, 2010), the input of which were obtained through the following five steps: (1) Tokenization: Tokenization breaks text into its atomic elements. In this study, each tweet was broken into a list of tokens, i.e. words originally separated by space or punctuations (Manning et al., 2008); (2) Stop word removal: we built a custom list of stop words by extending the nltk (Loper & Bird, 2002) stop words library with high occurrence tokens such as https, covid, coronavirus, etc., which are expected from most tweets, and don’t contribute to the characteristics of each topic; (3) Bigram transformation: although unigrams were used to build the bag of words (BOW) in our LDA model, we extended the BOW to include frequent bigrams such as New York, which were identified through a gensim bigram model (Řehůřek & Sojka, 2010) trained over our corpora; (4) Lemmatization: each token was lemmatized to its dictionary form, i.e. lemma (Manning et al., 2008); and (5) Bag of word encoding: using the lemmatized tokens, we can represent the corpora as two different vector spaces, with each vector component representing a lemma.

2.4. Measures

Popularity: For a tweet \( k \), we can represent its popularity as \( P_k = 1 + n_k^{\text{favorite}} + n_k^{\text{retweet}} \), where \( 1 + n_k^{\text{favorite}} + n_k^{\text{retweet}} \) represents the tweet itself, the number of favorites, and the number of retweets of tweet \( k \), respectively.

Topic size: For a topic \( j \), its topic size \( v_{j;k} \) in tweet \( k \) is measured by the percentage of tokens assigned to the topic by the LDA model. That is, \( v_{j;k} = \frac{n_{j;k}}{\sum_i n_{i;k}} \), where \( n_{j;k} \) denotes the number of tokens assigned to topic \( j \) in tweet \( k \). We note the property that all topic sizes in a single tweet normalizes to 1. That is, \( \sum_j v_{j;k} = 1 \). A total corpus topic size \( V_{j;k} \) is represented by the sum of individual topic sizes over all tweets, weighed by its popularity \( P_k \):

\[
V_j = \sum_k V_{j;k} P_k,
\]

where here and throughout the article, \( N \) represents the total number of tweets. The summation index \( k \) iterates through all tweets (1 to \( N \), if not otherwise specified.

A normalized corpus topic size is the total topic size over the corpus, normalized by the total number of tweets, favorite counts, and retweet counts. That is,
\[ V_j = \frac{V_j}{\sum_k P_k} \] (1)

**Sentiment score:** We used TextBlob (TextBlob, 2020), an open-source python library to generate the polarity score. For each tweet \( k \), a sentiment score \( S_k \) within the range from \(-1\) to \(1\), is included in the source data, where \(-1\) indicates the most negative, and \(1\) indicates the most positive sentiment, respectively. The sentiment \( S_j \) of a topic \( j \) can be represented as an average of \( s_k \) over all tweets, weighted by its topic size \( v_j \) as

\[ S_j = \frac{\sum_k v_j P_k s_k}{\sum_k v_j P_k} = \frac{\sum_k v_j P_k s_k}{\sum_k v_j P_k} \] (2)

**Significance:** The distribution of popularity \( P_k \) can be used to measure the significance of RQ 5. Tweets with positive and negative emotions can be extracted using thresholds \( S_+ \) and \( S_- \), respectively. T-statistics p-values can thus be measured on \( P_k | s_k \geq S_+ \) and \( P_k | s_k \leq S_- \).

**Time Variations:** to study the change of variables over the study period, we denote timestamps by \( t_i \), with \( t_0 \) and \( t_{\text{end}} \) representing March 20, and 9 August 2020, respectively. Except for \( t_{\text{end}} - t_{\text{end} - 1} \), all time stamps are separated by 1-week intervals. To study how the topic sizes change over time, we define \( V_j(t_i) \) as the normalized corpus topic size in time interval \( t_{i - 1} \) and \( t_i \). Therefore, c.f. Eq. 1, \( \sum_i V_j(t_i) = V_j \). In other words, \( V_j(t_i) \) represents the increment of topic size. We can also define the time-varying sentiment score \( S_j(t_i) \) as the weighted average similar to Eq. 2, but over tweets binned between two consecutive timestamps \( t_{i - 1} \) and \( t_i \). That is,

\[ S_j = \sum_{t_{i - 1} < t_i} v_j P_k s_k / \sum_{t_{i - 1} < t_i} v_j P_k \] (3)

3. Results

3.1. Topics of Twitter discourse on mask-wearing and vaccination

To find the most discussed topics on mask-wearing and vaccination on Twitter (RQ 1), optimal LDA models were developed on the two corresponding corpora, with coherence score \( C_v = 0.61 \) and 0.36, respectively. Normalized corpus topic sizes and high-frequency tokens for the two corpora are shown in **Table 2a** and **Table 2b**, where topic names were labeled based on the high-frequency tokens and sample where the corresponding topic predominates.

| Topic name       | Topic size | Topic sentiment | High-frequency tokens                  |
|------------------|------------|-----------------|----------------------------------------|
| Mask as new norm | 10.8%      | 0.11            | Wear, social distance, pandemic, protect, keep, stay |
| Face shield DIY  | 10.3%      | 0.10            | Bottle, empty, soda, recycle, billonsheild, challenge |
| Mask for love    | 18.8%      | 0.02            | Love, besafe, staysafe, trust, savelife, superhero |
| Ads/sales        | 46.8%      | 0.14            | Sale, washable, design, fashionable, price |
| Government criticism | 13.3% | 0.11            | Trump, operation, republican, idiot, leader, law, court |
The major topics and topic sizes on mask-wearing are (1) Mask as a new norm (10.8%), (2) Face shield DIY (10.3%), (3) Mask for love (18.8%), (4) Ads/sales (46.8%), and (5) Government criticism (13.3%). For vaccination, the major topics and corresponding sizes are (1) Science (25.8%), (2) Coping without vaccine (25.0%), (3) Immunity boost (14.9%), (4) Vaccine race (14.3%), and (5) Politics (20.0%).

To answer whether people’s interests in mask-wearing and vaccination topics evolved over time (RQ. 2) and whether the evolving pandemic and public policy changes influenced the topics (RQ. 6), we calculated $V_j(t)$ for both corpora throughout the study period. The results are shown in Figure 1 (a) and Figure 2 (a), respectively.

The topic size growth rate of all topics for mask-wearing underwent four significant increases around April 17th, May 15th, June 12th, and July 28th of 2020, respectively. Four critical events prior to the increases were identified, which were expected to raise public concerns on the control of COVID-19. For instance, on April 16th, The White House released broad guidelines for returning to work and reopening the economy (Edelman et al., 2020, April 16.); on May 12th, Dr. Anthony Fauci testified before the US Senate that the death toll of 80,000 was likely an underestimate (Shabad, 2020, May 12).

For vaccination, one major topic size growth was found for the Science and Coping without vaccine topics after July 7th, when the Trump administration gave formal notice of withdrawal from WHO (Smith & Perlmutter-Gumbiner, 2020, July 7). A minor increase was also found for the Science topic after May 5th, when researchers from Pfizer and New York University reported a coronavirus vaccine could be ready by September (Costello & Stelloh, 2020, May 5). The growth of other topics stayed rather steady over the study period.

### 3.2. Topic sentiment and popularity

To find out the valence of each of the detected topics about mask-wearing and vaccination (RQ. 3), the topic sentiment $S_j$ for the full study period (Eq. 2) is shown for mask-wearing and vaccination in Table 2 (a-b), respectively. Despite the expected concern on the pandemic, we found that topics about mask-wearing and vaccination were generally positive, except the Science topic on vaccination, which had a slightly negative sentiment of $-0.006$.

To answer how valence of mask-wearing and vaccination change over time (RQ. 4) and how the changes were influenced by the evolving pandemic and public policy (RQ. 7), sentiment scores for all topics in both corpora (Eq. 3) are calculated and shown in Figure 1 (b) and Figure 2 (b), respectively. Two major dips in sentiment, implying an increase in public concern on mask-wearing were found for all topics after April 17th and July 31st, when the White House announced

### Table 2b. Topic on vaccine by LDA model

| Topic name     | Topic size | Topic sentiment | High-frequency tokens                      |
|----------------|------------|-----------------|-------------------------------------------|
| Science        | 25.8%      | -0.006          | Pandemic, scientist, herd immunity, comparsis, immune_system |
| Coping without vaccine | 25.0%      | 0.005           | Deal, safe, distance, stay, work, home |
| Immunity boost | 14.9%      | 0.109           | Immune_system, boost, energy, vitamin, sleep |
| Vaccine race   | 14.3%      | 0.064           | Vaccine, race, test, headline, process |
| Politics       | 20.0%      | 0.127           | Mail, voter, rule, ballot |
its gating criteria to reopen the economy, and CDC called for reopening American schools, respectively. Note in Figure 1(a) that a topic size increase was observed at the same times.

For vaccination, a significant dip in sentiment was found for Science, Vaccine race, and Coping without vaccine topics after the single-day death rate in New York reached its peak on April 8th.
(Givetash, 2020, April 9), and after July 7th upon US official withdrawal from WHO (Smith & Perlmutter-Gumbiner, 2020, July 7). In the latter event, a dramatic increase in topic size was observed for Science and Coping without vaccine topics too (Figure 2 (a)). Sentiment rise was also found for Science, Vaccine race, and Coping without vaccine topics after Pfizer and NYU researchers indicated a vaccine from their research could be ready by September. The same event was also found to correlate to a significant topic size rise in Figure 2 (a). A rising trend in sentiment was found after July 15th for the Science and Coping without vaccine topics, after Moderna reported
safety for participants from their Phase I study (Clinical Trials Arena, 2020). The Immunity boost and Politics topic sentiments stayed roughly unchanged over the study period.

To measure the significance on whether users like to retweet and react to positive or negative content (RQ. 5), we assign a heuristic pair of the sentiment score thresholds ($S^+, S^-$) = ($-0.1$, $0.1$), for positive and negative emotions, respectively (see Sec. Measures). Twitter users retweeted and reacted to more positive content about mask-wearing ($t = 2.32, p = .02$). In contrast, negative content was more popular about vaccination ($t = -3.17, p = .0017$).

4. Discussion

Our LDA models identified five most discussed topics about mask-wearing on Twitter from March 20th to 9 August 2020 in the US: (1) wearing a mask as a new norm, (2) face shield DIY, (3) wearing a mask to protect others, (4) custom mask sale, and (5) criticism on government officials and mask related policies. In comparison, the five most discussed topics about vaccination were: (1) science behind vaccination, (2) coping the pandemic without vaccine (e.g., practicing social distance; working from home), (3) methods to boost immunity, (4) race of vaccine development among pharmaceutical companies, and (5) whether weak immunity can be used for a reason to mail-in votes. This study also explored the trends in people’s interests on different topics over time during the study period. On mask-wearing, we found that discussion was boosted in response to the policy changes, such as the White house’s decision to reopen the economy on April 15th, as well as the CDC’s call to reopen schools on July 23th. Mask-wearing related discussions also increased with negative news reported when the total infection cases in the United States passed two million on June 10 (AJMC, 2020), and when Dr. Fauci testified before the US Senate that the US death toll of 80,000 was likely an underestimate on May 12th. In comparison, discussion on vaccination increased significantly after the White house’s official withdrawal from WHO on July 6th, and after the news on Moderna’s success in its Phase I vaccine study on July 15th. These findings support the agenda setting theory (McCombs, 2004) that news media predicts the perceived salience of issues from the public, manifested as changes in topic sizes discussed on Twitter. In addition to the news coverage, government policy changes too, significantly drove the discussions on mask wearing and vaccination during the study period, which reflected the critical role the government plays in setting up public health agenda and influencing public reactions during a health crisis. In fact, with high uncertainty and relevance at both personal and social level, the agenda setting effects might have been amplified during the pandemic. Although knowledge about this new coronavirus is growing, the novelty and evolving nature of the pandemic still calls for a high need for orientation from the public. As people rely on media coverage and guidance from the government during the pandemic, the news media and the government play a critical role in shaping the public agenda. This study validated and emphasized the dominant agenda setting power of mainstream media and government policies during a public health crisis.

In addition to topic sizes, the sentiments on mask-wearing and vaccination were also found to be influenced by news about infectious rates, public policies, and health recommendations. On mask-wearing, after the White House’s decision to reopen the economy and CDC called to reopen American schools, the overall sentiment of the discussion became more negative, along with an increased discussion on mask-wearing. On vaccination, significant dips in sentiment were found for several topics (e.g. Science, Vaccine race) upon negative news about the pandemic, such as New York City reaching its single-day death peak and the US withdrawal from WHO. It is also noteworthy that as the sentiment of vaccination became more negative after the US withdrawal from WHO, more discussions about vaccination emerged, manifested as the dramatic increase in topic sizes for Science and Coping with vaccine. In contrast, positive updates on vaccine development were found to correlate with sentiment rise (e.g. Moderna reported safety for Phase I study). Therefore, it implies that people may generally be more engaged in the discussions about mask-wearing and vaccination when a high uncertainty level is expected and when the pandemic severity reaches certain milestones. During such time, people expressed more anxiety and other negative emotions in the tweeted messages, thus influencing the significant drop in sentiment score. It is astounding also, to find that although
the total death number was constantly increasing, the perceived severity of the pandemic from the public, manifested by the topic sizes and the average sentiment scores on mask-wearing and vaccination, was not increasing with the same steady trend, but rose only with noticeable media coverage. That is, the public perceived risk might have relied on the media reports. People need to be reminded by the media to understand the severity of the pandemic. This study supported the second level agenda setting effect such that news media is not only transferring issue salience to the public agenda, but also transferring attributes and the tone of these attributes (Coleman et al., 2008). Therefore, it is crucial for the media to exercise its power in agenda setting during the COVID-19 pandemic (Gever & Ezeah, 2020). As a counter example, a content analysis of six media outlets in Nigeria revealed that news stories about the COVID-19 virus only significantly increased after the first confirmed case in Nigeria, showing the media’s failure to send sufficient health warning messages before its national spread (Gever & Ezeah, 2020). Thus, it is also particularly important for the mainstream media to reflect on its media agenda during the COVID-19 crisis to pay attention to the marginalized and disadvantaged communities who had been hit the hardest during the pandemic (Milan & Treré, 2021).

We also aimed to study whether positive or negative emotional content were more transmissive on social media during the COVID-19 pandemic. Despite the pandemic, we found that the topics on mask-wearing and vaccination were generally positive, except the Science topic on vaccination, with a slightly negative sentiment of −0.006. In addition, users reacted differently on mask-wearing and vaccination to positive vs negative content. On topics about mask-wearing, more positive contents were retweeted and reacted to. This is consistent with results from several large scale content analysis that positive sentiment is more likely to be shared on social media (Al-Rawi, 2019; Berger & Milkman, 2012; Bollen et al., 2011; Tellis et al., 2019). In examining the content, much creativity was observed in tweets about mask-wearing, from custom mask making with bright colors and interesting patterns, to creative faceshield DIY, using bottles and other recycled material. Selfies were also posted as a sign of caring for others and being responsible. This finding indicates that despite the pandemic, people have the same motivations for posting and sharing content on social media as usual. Self-enhancement is still one of the most important motivations for posting and retweeting information related to mask-wearing (Tellis et al., 2019). Therefore, public health campaigns on promoting mask-wearing can focus on the positive and creative aspects to overcome the stigma and personal discomfort. Encouraging selfie posting can also promote a sense of participation and create a desired social norm on social media.

On topics about vaccination, however, negative tweets were retweeted and reacted to more frequently. This might be due to the high level of uncertainty and anxiety related to vaccination. Negative high-arousal emotions such as anxiety are among the most effective triggers to promote sharing motivation and have been found to spread quickly, especially on social media platforms (Berger & Milkman, 2012; Fan et al., 2014). It is worth noting that these emotions could also be the fuel for conspiracy theory. Although not identified as most discussed topics, a few conspiracy theories against vaccination (e.g. man-made COVID-19 virus for vaccine profit) have already emerged online during our study period. Therefore, the goal of public campaigns to promote vaccination can also be a battle against misinformation and conspiracies.

Our mixed findings echoed the importance of studying emotion and content virality under different contexts. The analysis among sentiment, topic size, and social media reactions contributed to the literature of social transmission of information on Twitter, particularly during public health crises. In the specific context of pandemic, people share more positive content about preventative measures with low uncertainty, such as mask-wearing, but more negative content about preventative measures with high uncertainty, such as vaccination. This finding shed lights on a potential new angle to study social transmission of information on social media.
5. Implications for health campaigns
This study analyzed US-based tweets from March to August, 2020 to understand the evolution of topics and sentiment about mask-wearing and vaccination. The findings suggested plausible transfer of salience of topics from news media coverage and government policy changes to Twitter discourse of both preventative measures of mask-wearing and vaccination, which was manifested as changes in topic sizes discussed on Twitter, and people’s emotional responses in such discussions. This underscored the importance of timing of health campaigns to promote preventative measures. Public health practitioners should utilize critical events to promote preventive measures such as mask-wearing and vaccination. More specifically, government health sectors should not only utilize news conferences to coordinate with news media to provide critical health information and policy updates, but also employ a more holistic approach to prepare health campaigns with timelines aligned with such news media reports. This requires coordination across different sectors and multiple levels within the healthcare systems. Such timing could be more accurate with the analyses of social media insights. When the size of a topic significantly increases on social media, it could be the optimal timing to launch a social media health campaign for the related topic to reach its maximum impact. In addition, analyses of social media sentiment after critical events could also help understand people’s emotional responses to such events and thus help optimize the content for most effective health campaigns accordingly.

The findings also underscored the importance to investigate sentiment valence and content vitality for different preventative measures. The analysis showed that tweets about mask-wearing and vaccination were generally positive, except for the science topic on vaccination. Also, people retweeted more positive content about mask-wearing but more negative content about vaccination. Understanding social media sentiment for different topics regarding preventive measures would provide guidance on how to promote compliance to the NPI measures and vaccination. For a preventive measure associated with negative sentiment, it is important to examine its reason so that a health campaign can target these negative emotions to address the underlying concerns. We suggest that for preventative measures such as mask-wearing with less uncertainty, public health campaigns can promote positivity and creativity to encourage participation, thus creating a desired social norm on social media to combat stigma and personal discomfort. For preventative measures such as vaccination with relatively higher uncertainty, especially at the vaccine development stage, frequent updates with more detailed information can be used to reduce anxiety, and to confront misinformation and conspiracies.

6. Limitations and future studies
This study has several limitations. First, the source data has not captured the tweets beyond the keywords used for scraping the tweets. Although this extensive dataset receives daily updates, some negative hashtags such as #NoMask was not included as keywords. In addition, social desirability might also prohibit posting opinions against the mainstream recommendations.

Secondly, although sentiment analysis at scale provides intriguing features to study content from online social networks, no causal relationship can be established based on the methodology. Therefore, we might know which sentiments are the most spread, but may not conclude on which sentiment leads to the sharing intention without testing the intention itself.

Future studies can focus on capturing data on misinformation during the pandemic and also further exploring the role of discrete emotions on social sharing during the pandemic. The sentiment analysis of the present study focuses on valence and its relationship with social transmission. Besides emotional valence, research also suggests the importance of contextual information when considering the link between affection and content virality. Despite the collective reflection of the positivity on social media, social transmission of emotional content was driven not only by emotion valence but also the activation they evoke (Berger & Milkman, 2012; Fan et al., 2014). It would be interesting for future studies to study emotions and its relationship with social transmission.
Funding
The authors received no direct funding for this research.

Author details
Yi (Jasmine) Wang
E-mail: jasmine.wang@louisville.edu
ORCID ID: http://orcid.org/0000-0003-2293-3664
Mouli Shi
E-mail: shimoli@alum.mit.edu
Jueman (Mandy) Zhang
E-mail: jzhang19@nyit.edu

References
Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infowveillance Study. Journal of Medical Internet Research, 22(4), e19016. https://doi.org/10.2196/19016
AJMC (2020, June 11). What we’re reading: Black youth mental health tested; COVID-19 cases top 2 million; nursing homes seizing stimulus checks. https://www.ajmc.com/news/black-youth-mental-health-tested-covid19-cases-top-2-million
Akpinar, E., & Berger, J. (2017). Valuable virality. Journal of Marketing Research, 54(2), 318–330. https://doi.org/10.1509/jmr.13.0350
Al-Rawi, A. (2019). Viral news on social media. Digital Journalism, 7(1), 63–79. https://doi.org/10.1080/21670811.2017.1387062
Clinical Trials Arena, C. T. (2020, July). Moderna says Covid-19 vaccine candidate safe in Phase 1 study. https://www.clinicaltrialsarena.com/news/moderna-covid-vaccine-candidate-data/
Bontempo, E. O. C., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. Psychological Assessment, 21(1), 79–88. https://doi.org/10.1037/a0014643
Berger, J., & Milkman, K. (2012). What makes online content viral? Journal of Marketing Research, 49(2), 192–205. https://doi.org/10.1509/jmr.10.0353
Bhat, M., Qadri, M., Beg, N., Ul, A., Kundroo, M., Ahanger, N., & Agarwal, B. (2020). Sentiment analysis of social media response on the Covid19 outbreak. Brain, Behavior, and Immunity, 87, 136–137. https://doi.org/10.1016/j.bbi.2020.05.006
Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022. http://doi.org/10.1162/jmlr.2003.3.4-5.993
Bollen, J., Gonçalves, B., Ruan, G., & Mao, H. (2011). E: Intelligent, Not Evil: The Complexity of Social Network Inference. Artificial Life, 17(3), 237–251. https://doi.org/10.1162/artl_a_00034
Cobling, M. L., Turner, J. W., Hurtado-de-Mendoza, A., Zhang, Y., Jiang, X., Drago, F., & Sheppard, V. B. (2018). Sentiment analysis of an online breast cancer support group: Communicating about Tamoxifen. Health Communication, 33(9), 1158–1165. https://doi.org/10.1080/10410236.2017.1339370
Cambria, E., Schuller, B., Xiao, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15–21. https://doi.org/10.1109/MIS.2013.30
Carley, K., Molik, M., Landwehr, P., Pfeffer, J., & Kowalchuck, M. (2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. Safety Science, 90, 48–61. https://doi.org/10.1016/j.ssci.2016.04.002
CDC (2020, June 12). Public Attitudes, Behaviors, and Beliefs Related to COVID-19, Stay-at-Home Orders, Nonessential Business Closures, and Public Health Guidance — United States, New York City, and Los Angeles, May 5–12, 2020. https://www.cdc.gov/mmwr/volumes/69/wr/mm6926e1.htm
Chandrasekar, R., Mehta, V., Vollkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. Journal of Medical Internet Research, 22(10), e22624. https://doi.org/10.2196/22624
Chew, C., & Eysenbach, G. (2013). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE, 8(11), e11418. https://doi.org/10.1371/journal.pone.0011418
Coleman, R., McCombs, M., Shaw, D., & Weaver, D. (2008). Agenda setting. In K. Wahl-Jorgensen & T. Hesitens (Eds.), The Handbook of Journalism Studies (pp. 147–166). Routledge. https://doi.org/10.4324/9780203877685
Conway-Silva, B. A., Filer, C. R., Kensi, K., & Tsetsi, E. (2018). Reassessing Twitter’s agenda-building power: An analysis of intermediaria media-setting effects during the 2016 presidential primary season. Social Science Computer Review, 36(4), 469–483. https://doi.org/10.1177/0894439317715540
Costello, T., & Stelich, T. (2020, May 5). Pfizer, NYU working on innovative coronavirus vaccine that could be ready by end of summer. NBC News. https://www.nbcnews.com/nightly-news/pfizer-nyu-working-innovative-coronavirus-vaccine-could-be-ready-end-n1200776
Doogan, C., Buntine, W., Linger, H., & Brunt, S. (2020). Public perceptions and attitudes toward COVID-19 nonpharmaceutical interventions across six countries: A topic modeling analysis of Twitter data. Journal of Medical Internet Research, 22(9), e21419. https://doi.org/10.2196/21419
Edelman, A., Alba, M., Nichols, H., & Alexander, P. (2020, April 16). Trump unveils three-phase plan for states to reopen amid coronavirus pandemic. NBC News. https://www.nbcnews.com/politics/donald-trump/trump-announce-guidelines-reopening-u-s-amid-coronavirus-pandemic-n1185666
Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger is more influential than joy: Sentiment correlation in weibo. PLoS ONE, 9(10), e101184. https://doi.org/10.1371/journal.pone.0101184
Geoocoder, R. (2020). https://github.com/thompsonman/reverse-geoocoder
Gever, V. C., & Ezechie, G. (2020). The media and health education: Did Nigerian media provide sufficient warning messages on coronavirus disease? Health Education Research, 35(1), 460–470. https://doi.org/10.1093/her/cya0020
Givetash, L. (2020, April 9). Coronavirus pandemic: N.Y. now has more reported cases than any country. NBC News. https://www.nbcnews.com/news/world/coronavirus-pandemic-ny-now-has-more-reported-cases-any-country-n1179836
Wang et al., Cogent Social Sciences (2021), 7: 1959728
https://doi.org/10.1080/23311886.2021.1959728

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., … Verlegh, P. (2005). The firm’s management of social interactions. Marketing Letters, 16(3–4), 415–428. https://doi.org/10.1007/s11002-005-5902-4

Goodnough, A. (2020, July 24). C.D.C. Calls on Schools to Reopen, Downplaying Health Risks. The New York Times. https://www.nytimes.com/2020/07/24/health/cdc-schools-coronavirus.html

Günther, E., & Domahidi, E. (2017). What communication scholars write about: An analysis of 80 years of research in high-impact journals. International Journal of Communication, 11, 3051–3071.

Hansen, L. K., Arvidsson, A., Nielsen, F. A., Colleoni, E., & Etter, M. (2011). Good friends, bad news - Affect and virality in Twitter. In J. J. Park, L. T. Yang, & C. Lee (Eds.), Future Information Technology. Communications in Computer and Information Science (pp. 185). Springer. https://doi.org/10.1007/978-3-642-23309-9_5

Hatchett, R. J., Mecher, C. E., & Lipsitch, M. (2007). Public health interventions and epidemic intensity during the 1918 influenza pandemic. Proceedings of the National Academy of Sciences of the United States of America, 104(18), 7562–7587. https://doi.org/10.1073/pnas.0610941104

HHS (n.d.). COVID-19 and flu public education campaign. July 3, 2020. https://www.hhs.gov/coronavirus/education-campaign/index.html

Igelnik, R. (2020, June 23). Most Americans Say They Regularly Wore a Mask in Stores in the Past Month; Fewer See Others Doing it. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/06/23/most-americans-say-they-regularly-wore-a-mask-in-stores-in-the-past-month-fewer-see-others-doing-it/

John Hopkins Coronavirus Resource Center. (n.d.) Covid-19 Map. July 3, 2021. https://coronavirus.jhu.edu/map.html

Keim-Malpass, J., Mitchell, E. M., Sun, E., & Kennedy, C. (2017). Using twitter to understand public perceptions regarding the #HPV Vaccine: opportunities for public health nurses to engage in social marketing. Public Health Nursing, 34(4), 316–323. https://doi.org/10.1111/phn.12318

Kim, H. S. (2015). Attracting views and going viral: How message features and news-sharing channels affect health news diffusion. Journal of Communication, 65(3), 512–534. https://doi.org/10.1111/jcom.12160

Kwok, S. W. H., Vadde, S. K., & Wang, G. (2021). Tweet topics and sentiments relating to COVID-19 vaccination among Australian twitter users: Machine learning analysis. Journal of Medical Internet Research, 23(5), e26953. https://doi.org/10.2196/26953

Lamsal, R. (2020). Coronavirus (covid-19) tweets dataset. IEEE Dataport. https://eprint.iacr.org/open-access/coronavirus-covid-19-tweets-dataset

Li, T., Liu, Y., Li, M., Qian, X., & Dai, S. Y. (2020). Mask or no mask for COVID-19: A public health and market study. PLoS One, 15(8), e0237691. https://doi.org/10.1371/journal.pone.0237691

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, Q., Zheng, Z., Zheng, J., Chen, Q., Liu, G., Chen, S., … Ming, W. K. (2020). Health communication through news media during the early stage of the covid-19 outbreak in China: Digital topic modeling approach. Journal of Medical Internet Research, 22(4), e19118. https://doi.org/10.2196/19118

Live Twitter Sentiment (2020). https://live.lamsal.com/npl/

Loper, E., & Bird, S. (2002). NLTK. In Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics (Vol. 1, pp. 63–70). Association for Computational Linguistics (ACL). https://doi.org/10.3115/1118108.1118117

Manning, C. D., Raghavan, P., & Schutze, H. (2008). Introduction to Information Retrieval. Cambridge University Press. https://doi.org/10.1017/cbo9780511809071

McCombs, M. E. (2004). Setting the agenda: The mass media and public opinion. Addien, MA: Polity.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. Public Opinion Quarterly, 36(2), 176–187. https://doi.org/10.1086/267990

Meraz, S. (2009). Is there an elite hold? traditional media to social media agenda setting influence in blog networks. Journal of Computer-Mediated Communication, 14(3), 682–707. https://doi.org/10.1111/j.1083-6101.2009.01458.x

Milan, S., & Trére, E. (2021). A widening data divide: COVID-19 and the global south. In S. Milan, É. Trére, & S. Masiero (Eds.), Pandemic Inequalities, Policies and Resistance in the Digitalized Society (pp. 24–28). The Institute of Network Cultures, Nikolakou, A., & King, K. W. (2018). Viral video ads: Emotional triggers and social media virality. Psychology and Marketing, 35(10), 715–726. https://doi.org/10.1002/mar.21229

O’Keefe, S. M. (2020, Aug. 7). One in Three Americans Would Not Get COVID-19 Vaccine. Gallup News. https://news.gallup.com/poll/317018/one-three-americans-not-covid-vaccine.aspx

Odlum, M., & Yoon, S. (2015). What can we learn about the Ebola outbreak from tweets? American Journal of Infection Control, 43(6), 563–571. https://doi.org/10.1016/j.ajic.2015.02.023

Ogata Jones, K., Denham, B. E., & Springston, J. K. (2006). Effects of mass and interpersonal communication on breast cancer screening: Advancing agenda-setting theory in health contexts. Journal of Applied Communication Research, 34(1), 94–113. https://doi.org/10.1080/00909880500420424

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., … Wu, T. (2020). Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA, 323(19), 1915–1921. https://doi.org/10.1001/jama.2020.613

Patel, M. D., Rosenstrom, E., Ivy, J. S., Mayorga, E. M., Keskinokac, P., Boyce, R. M., Lich, K. H., Smith, R. L., Johnson, K. T., & Swan, J. L. (2020). The joint impact of COVID-19 vaccination and non-pharmaceutical interventions on infections, hospitalizations, and mortality: An agent-based simulation. medRxiv: the preprint server for health sciences. 2020.12.30.20248888. https://doi.org/10.1101/2020.12.30.20248888

Paul, M. J., & Dredze, M. (2011). You are what you tweet: Analyzing Twitter for public health. Proceedings of the fifth international conference on Weblogs and social media (ICWSM), Barcelona, 265–272. 71.

Pratt, C. B., Ha, L., & Pratt, C. A. (2002). Setting the Public Health Agenda on Major Diseases in Sub-Saharan Africa: African Popular Magazines and Medical Journals, 1981–1997. Journal of Communication, 52(4), 889-905. https://doi.org/10.1111/j.1460-2466.2002.tb02379.x

Rěhůrek, R., & Sojka, P. (2010). Gensim—Statistical Semantics in Python. EuroScipy 2011, Paris.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In WSDM 2015 - Proceedings of the 8th ACM International Conference on
on Web Search and Data Mining (pp. 399–408). Association for Computing Machinery, Inc. https://doi.org/10.1145/2684822.2685324

Shabad, R. (2020, May 12). Fauci warns ‘little spikes’ of coronavirus could turn into outbreaks if states reopen too soon. NBC News. https://www.nbcnews.com/politics/congress/fauci-set-testify-congress-us-coronavirus-response-reopening-plans-n1205051

Smith, A., & Perlmutter-Gumbiner, E. (2020, July 7). Trump administration gives formal notice of withdrawal from WHO. NBC News. https://www.nbcnews.com/politics/donald-trump/trump-administration-gives-formal-notice-withdrawal-who-n1233100

Snyder, L. B., & Hamilton, M. A. (2002). Health evidence–Meta-analysis of U.S. health campaign effects on behavior: Emphasize enforcement, exposure, and new information, and beware the secular trend. In R. Hornik (Ed.), Public Health Communications: Evidence for Behavior Change (pp. 357–384). Lawrence Earbaum.

So, J., Prestin, A., Lee, L., Wang, Y., Yen, J., & Chou, W. Y. S. (2016). What do people like to “share” about obesity? A content analysis of frequent retweets about obesity on Twitter. Health Communication, 31(2), 193–206. https://doi.org/10.1080/10410236.2016.940675

Stieglitz, S., & Dong-Xuan, L. (2013). Emotions and information diffusion in social media - Sentiment of microblogs and sharing behavior. Journal of Management Information Systems, 29(4), 217–248. https://doi.org/10.2753/MIS0742-1222290408

Su, Y., & Borah, P. (2019). Who is the agenda setter? Examining the intermedia agenda-setting effect between Twitter and newspapers. Journal of Information Technology & Politics, 16(3), 236–249. https://doi.org/10.1080/19331681.2019.1641451

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence. Journal of Marketing, 83(4), 1–20. https://doi.org/10.1177/0022242919841034

TextBlob, (2020). TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/

Twarc, (2020). https://github.com/DocNow/twarc

Valenzuela, S., Puente, S., & Flores, P. (2017). Comparing disaster news on Twitter and television: An intermedia agenda setting perspective. Journal of Broadcasting & Electronic Media, 61(4), 615–637. https://doi.org/10.1080/08838151.2017.1344673

WHO (n.d.). Coronavirus disease (COVID-19) pandemic. July 3, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses

Xu, Z. (2020). How emergency managers engage Twitter users during disasters. Online Information Review, 44 (4), 933–950. https://doi.org/10.1108/OIR-08-2019-0275

Zucker, H. G. (1978). The variable nature of news media influence. Communication Yearbook, 2(1), 225–240. https://doi.org/10.1080/23808985.1978.11923728