Public Discourse Against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data

Mohammad Al-Ramahi, Ahmed El Noshokaty, Omar El-Gayar, Tareq Nasralah, Abdullah Wahbeh

Submitted to: JMIR Public Health and Surveillance
on: December 25, 2020

Table of Contents

Public Discourse Against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data

Mohammad Al-Ramahi[1] DSC; Ahmed El Noshokaty[2] DSC; Omar El-Gayar[3] PhD; Tareq Nasralah[4] PhD; Abdullah Wahbeh[5] DSC

[1]Texas A&M University-San Antonio San Antonio US
[2]Northern Michigan University Marquette US
[3]Dakota State University Madison US
[4]Supply Chain and Information Management Group D'Amore-McKim School of Business Northeastern University Boston US
[5]Slippery Rock University of Pennsylvania Slippery Rock US

Corresponding Author:
Tareq Nasralah PhD
Supply Chain and Information Management Group
D'Amore-McKim School of Business
Northeastern University
360 Huntington Ave
Boston
US

Abstract

Background: Despite the presence of scientific evidence supporting the importance of wearing masks to curtail the widespread of the COVID-19 virus, wearing masks has stirred up a significant debate particularly on social media.

Objective: To investigate the topics associated with the public discourse against wearing masks. Further, we study the relationship between the anti-mask discourse on social media and the number of new COVID-19 cases.

Methods: Using hashtags against wearing masks, we collected a total of 51,170 English tweets between January 1st, 2020 and October 27th, 2020. We used machine learning techniques to analyze the data collected. We investigate the relationship between the volume of tweets that are against mask-wearing and the daily volume of new COVID-19 cases using the Pearson Correlation between the two time series.

Results: The results and analysis showed that social media could help identify important insights related to wearing masks. The results of topic mining identified 10 categories/themes of user concerns dominated by 1) constitutional rights and freedom of choice followed by 2) conspiracy theory, population control and big pharma, and 3) Fake news, fake numbers, fake pandemic. Combined, these three categories represent almost 65% of the volume of tweets against masks. The relationship between the volume of tweets against wearing masks and the reported new COVID-19 cases depicts a strong correlation where the rise in the volume of negative tweets is leading the rise in the number of new cases by nine days.

Conclusions: The findings demonstrated the potential of mining social media for understanding the public discourse about public health issues such as wearing masks during the COVID-19 pandemic. The results emphasize the relationship between the discourse on social media and the potential impact on real events like changing the course of the pandemic. Policy makers are advised to proactively address public perception and work on shaping this perception through raising awareness, debunking negative sentiments, and prioritizing early policy intervention toward the most prevalent topics.

(JMIR Preprints 25/12/2020:26780)
DOI: https://doi.org/10.2196/preprints.26780

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
   Please make my preprint PDF available to anyone at any time (recommended).
   Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
✓

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Public Discourse Against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data

Abstract

Background: Despite the presence of scientific evidences supporting the importance of wearing masks to curtail the widespread of the COVID-19 virus, wearing masks has stirred up a significant debate particularly on social media.

Objective: To investigate the topics associated with the public discourse against wearing masks in the United States. Further, we studied the relationship between the anti-mask discourse on social media and the number of new COVID-19 cases.

Methodology: Using hashtags against wearing masks, we collected a total of 51,170 English tweets between January 1st, 2020 and October 27th, 2020. We used machine learning techniques to analyze the data collected. We investigated the relationship between the volume of tweets that are against mask-wearing and the daily volume of new COVID-19 cases using the Pearson Correlation between the two-time series.

Results: The results and analysis showed that social media could help identify important insights related to wearing masks. The results of topic mining identified 10 categories/themes of user concerns dominated by 1) constitutional rights and freedom of choice followed by 2) conspiracy theory, population control and big pharma, and 3) Fake news, fake numbers, fake pandemic. Combined, these three categories represent almost 65% of the volume of tweets against masks. The relationship between the volume of tweets against wearing masks and the reported new COVID-19 cases depicted a strong correlation where the rise in the volume of negative tweets is leading the rise in the number of new cases by nine days.

Conclusion: The findings demonstrated the potential of mining social media for understanding the public discourse about public health issues such as wearing masks during the COVID-19 pandemic. The results emphasized the relationship between the discourse on social media and the potential impact on real events like changing the course of the pandemic. Policy makers are advised to proactively address public perception and work on shaping this perception through raising awareness, debunking negative sentiments, and prioritizing early policy intervention toward the most prevalent topics.

Keywords: pandemic, coronavirus, masks, social medial, opinion analysis.

Introduction

The corona-virus disease 2019 (COVID-19) is an infection caused by the new coronavirus (SARS-CoV-2) that could cause acute repository syndrome [1]. As of December 26th, 2020; the COVID-19 affected 192 countries around the word, with a total of 80,416,535 reported cases, and a total of 1,757,888 deaths [2]. The World Health Organization (WHO), The Center for Disease Control (CDC), and other leading public health organizations have outlined several guidelines to heed off the COVID-19 pandemic. These guidelines have also been reported in recent scientific studies regarding the spread of the COVID-19 virus. The success of initiatives aimed at opening the national and regional (state) economies ultimately relies on public awareness and acceptance of these guidelines for limiting the transmission of the COVID-19 virus. Among these guidelines is the importance of wearing masks.

Existing studies [e.g., 3, 4] showed that masks could have a substantial impact on virus transmission, where wearing masks might significantly decrease the number of new COVID-19 cases. It is shown that wearing a mask was more effective than just hand washing [5]. The prevalence of evidence showed that mask-wearing diminishes disease spread by reducing the transmission probability per contact. Public mask-wearing is most effective at stopping the spread of the virus when compliance is high [6] and presents a rational way to implement as a non-pharmaceutical intervention (NPI) to fight COVID-19 [7]. Wearing a face mask can be effectively combined with social distancing to flatten the epidemic curve [7]. For individuals, an effective method of adequate isolation is to wear a mask [8]. Ma, Shan [9] found that N95 masks, medical masks, and even homemade masks could block at least 90% of the virus in aerosols. Wang, Pan [10] found that the necessity of wearing masks by the public during the COVID-19 pandemic has been under-emphasized. Nevertheless, and despite its importance supported by scientific evidence, wearing masks has stirred up a significant debate, particularly in the United States.

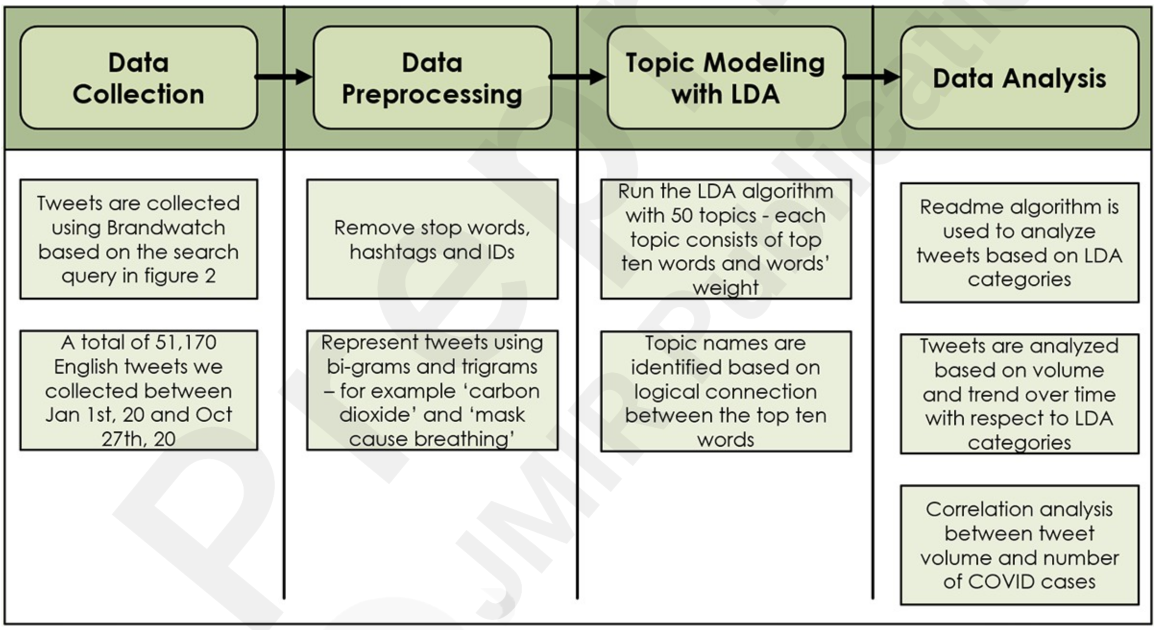
With millions of people forced out of public spaces, many conversations about such phenomena now take place on social media [11]. Popular social media platforms, including Twitter, enable new channels for users to share information and their experiences [12]. These platforms have provided efficient methods of information access for health surveillance and social intelligence [13-15], and have a growing popularity for sharing and debating scientific information [16-18]. Using Twitter as a data source, several studies have demonstrated the potential to identify the public's reactions to a variety of public health crises. Examples include the opioid epidemic [19, 20], marijuana [21-23], and vaping [24]. However, there is limited studies that examined the public discourse against masks on social media and its potential relation to the rise of COVID-19 cases.

With plenty of evidence supporting the effectiveness of masks in mitigating the spread of COVID-19, the vigorous public debate about masks is still ongoing [25] . Accordingly, in this research we aim to provide insights into factors and topics encompassing the ongoing (and sometimes contentious) debate surrounding mask-wearing. Specifically, our research objective is to investigate the topics associated with the public discourse against wearing masks. The study also analyzed trends overtime for each topics with a particular emphasis on the relative volume for each topic, and the spikes in volume. Further, we studied the relationship between the anti-mask discourse on social media and the number of new COVID-19 cases. The time lagged cross correlation (TLCC) is used to identify directionality between two signals; volume of tweets and covid-19 case; to determine which signal occurs first by looking at cross correlations, where peak correlation may have a different offset if one signal leads another. The analysis provided insights into the potential relationship between the cyber world represented by activities on social media and the physical world reflected by individuals' actions and possibly reflected in increased infection rates. Such understanding is needed as governments and public health officials grapple with reopening the economies (and keeping them open) in a manner that does not aggravate the COVID-19 pandemic as a public health crisis of epic proportions.

Research Method

Figure 1 shows the methodology adopted in this study for mining social media. It started with data collection. The researchers agreed on a time period of interest to collect data and keywords (i.e., hashtags) used to search for online tweets. Second, the tweets collected were preprocessed by removing stop words, keywords with IDs, and Hashtags and represented using bi- and tri-grams. Third, a topic modeling technique, the Latent Dirichlet Allocation (LDA) algorithm [26], was used to analyze the preprocessed tweets to identify the prominent topics/categories in the posts. Last but not least, a social media analytics tool, Brandwatch, was used to analyze the frequency and track the predefined categories' volume over time. Brandwatch, a social media analytics company, employs unsupervised and supervised machine learning techniques and a text analysis model developed by Hopkins and King [27].

Figure 1. Methodology for Mining Social Media.



Data Collection

Our target social media platform for data collection was Twitter as a microblogging platform. Initially, we identified all hashtags against wearing masks who were actively discussing masks on Twitter. Using Brandwatch with the search query shown in Figure 2, we extracted all tweets for the identified hashtags between January 1st, 2020 and October 27th, 2020. A total of 51,170 English tweets were collected. The hashtags were identified in the literature [28] as well as exploring similar trending hashtags used against wearing masks at hashtags.org and hashtagify.me. A key advantage of using a social media analytics platform such as Brandwatch is that it provides access to the "Twitter firehose" (i.e., every public tweet ever posted on Twitter in any language and from any geographic location that meets the search criteria).

Figure 2. Hashtags and Search Query Used for Data Collection.

```
(#maskoffamerica  OR #NoMask* OR #takeoffthemask OR #maskburning OR #BurnYourMask OR
#burnyourmaskchallenge OR #masksdontwork OR #nomaskonme OR #NoMasksOnMe OR #nomaskselfie OR
#maskhoax OR #maskshoax OR #NoMasksEVER  OR #NoMoreMasks OR #IWillNotWearAMask OR #MaskOff*
OR #MasksOff* OR #facefreedom  OR #masksmakemesweaty OR #MasksAreDangerous OR #TakeTheMaskOff
OR #Stopforcingmasksonme OR #takeoffyourmask* OR #refusemask* OR  #NeverMasker OR
#StopWearingMask* OR #StopWearingTheDamnMasks OR #MasksdontMatter OR #stopmasking OR
#stopthestupidmask OR #maskingchildrenischildabuse OR #MomsAgainstMasks OR #MasksRUnhealthy OR
#SheepWearMasks OR #MasksAreForSheep OR  #RefuseToWearMasks OR #MasksAreMurderingMe)
AND - (RT OR http OR https)
```

For comparing the volume of tweets against wearing masks and the number of COVID-19 cases, we collected a time series of the daily number of reported COVID-19 new cases in the US from January to October from John Hopkins University [29]. We have also collected the reported new COVID-19 cases daily from 22nd January to 27th October in the US.

In acquiring data from Twitter, the current study takes into consideration all the common regulatory concerns that arise with social media research. Specifically, the study conforms with federal regulations on research about human subjects by using only public information that requires no interaction with the poster [30]. Further, the use of Brandwatch ensured that the study conformed with all the common ethical questions raised when performing web mining [31].

Data Pre-processing

We excluded retweets and addresses to focus on personal opinions/statements. First, collected tweets were preprocessed by removing stop words as well as removing keywords with IDs and Hashtags. Second, tweets were represented using uni-gram, bi-gram and tri-grams such as 'results', 'lab results', and 'check test results'. The reason word-level n-grams features were selected to represent tweets over bag-of-words (i.e., single words) features is that the latter has two major drawbacks: 1) they lose the ordering of the words and 2) they ignore semantics of the words [32, 33].

Data Analysis Using the LDA Algorithm (Unsupervised Learning)

To discover the abstract “topics” that occur in the collected posts, we ran a topic mining model, specifically the Latent Dirichlet Allocation (LDA) algorithm, with 50 topics. Given a set of documents, $D = \{d_1, d_2, \ldots, d_n\}$; a number of topics, $T = \{t_1, t_2, \ldots, t_m\}$; and a number of words in each topic, $W = \{w_1, w_2, \ldots, w_k\}$; the LDA algorithm generates the followings:

- A $D \times T$ matrix with $n \times m$ size, where the weight $w_{i,j}$ is the association between a document d_i and a topic t_j. [34]
- A $T \times W$ matrix with $m \times k$ size, where the weight $w_{i,j}$ is the association between the topic t_i and a word w_j. [34]

The corresponding reproductive process is shown below [34, 35], :

(1) For each topic $t \in \{1, \ldots, m\}$,

(a) generate a probability distribution over words

$\beta t \sim \text{Dirichlet}(\eta)$.

(2) For each document d,

(a) generate a vector of the topic probability distribution

$\theta d \sim \text{Dirichlet}(\alpha)$.

(b) For each word w_i in document d,

(i) generate a topic assignment

$z_i \sim \text{Multinomial}(\theta d)$;

(ii) generate a word $w_n \sim Multinomial(\beta_{zi})$

βt is the word distribution for topic t, and θd is the topic distribution for document d. The notations η and α are model parameters.

Topic models are statistical-based models for uncovering the main themes (i.e., set of topics) that depict a large and unstructured collection of documents. Topic models make it possible to summarize textual data at a scale that is not possible to be tackled by human annotation. In this study, we chose the Latent Dirichlet Allocation (LDA) algorithm [26] due to its conceptual advantage over other latent topic models [34-37].

The 50 topics from the LDA were labeled by the first author and validated by the second author. The identified topics were further analyzed and grouped into ten representative categories. The grouping was done based on semantic similarities between the topics identified. For example, the following topics “build herd immunity”, “herd Immunity”, “build immune system” could be grouped into in one main topic, namely “herd immunity and dependency on the immune system”. Overall, we have discovered and ended up with ten categories.

Analysis of Tweets Using Categories Obtained (Supervised Learning)

Brandwatch employs the ReadMe supervised algorithm developed by Daniel Hopkins and Gary King [27]. The ReadMe algorithm is particularly suited when the objective is to know the proportion of the population of posts that fit in specific categories. Rather than calculating these proportions based on the categorization of individual posts, ReadMe gives approximately unbiased estimates of category proportions even when the optimal classifier performs poorly [27].

The ReadMe algorithm requires the researcher to hand-code a ‘training set” of documents into a set of predefined categories. In this study, the tweets represent the set of documents and the predefined categories are obtained from the LDA algorithms. The authors hand-coded a total of twenty tweets into each predefined category obtained from the LDA, and then ran the ReadMe algorithm iteratively on the remaining posts, ensuring that the examples clearly outline each category. Then, based on the training phase, the algorithm will build a model that can automatically assign the remaining tweet to categories and obtain the total number of tweets in each category. Trends of tweets volumes over time is automatically generated by Brandwatch.

Analyzing the Relationship Between the Tweet Volume and the Number of COVID Cases

To analyze the relationship between the volume of tweets against mask-wearing and the daily volume of new COVID-19 cases, we plotted two-time series over the time span from January to October 2020 and calculated the Pearson Correlation. The Pearson Correlation measures how two continuous waves co-vary over time and indicate the linear relationship as a number between -1 (negatively correlated) to 0 (not correlated) to 1 (perfectly correlated) [38]. The correlation is a snapshot measure of global synchrony. Although the Pearson correlation provides a very simple way to compute both global and local synchrony, it does not provide insights into signal dynamics such as which signal occurs first which can be measured via cross correlations. Time lagged cross correlation (TLCC) can identify directionality between two signals such as a leader-follower relationship. We can get a sense of which signal occurs first by looking at cross correlations. TLCC is measured by incrementally shifting one time series vector and repeatedly calculating the correlation between two signals. If the peak correlation is at the center (offset=0), this indicates that the two-time series are perfectly synchronized at that time. However, the peak correlation may have a different offset if one signal leads another [39]. To analyze the relationship between the two-time series, volume of tweets against mask-wearing and daily volume of new COVID-19 cases, we calculated the Pearson correlation and TLCC in Python using the SciPy package.

Results

Tweet Distribution and Categories

A total of 51,170 tweets were analyzed with respect to categories identified from the LDA model. These categories were mainly related to (ordered per their frequency in posts) 1) constitutional rights and freedom of choice, 2) conspiracy theory, population control and big pharma, 3) fake news, fake numbers, fake pandemic, and lies, 4) unhealthy, low oxygen, Carbon dioxide, and lung infections, weaken immune system, 5) political, fear, and control people, 6) masks ineffective and cannot block tiny particles, 7) mental health and suicide, 8) Herd immunity and dependency on the immune system, 9) child abuse and dehumanization, 10) virus related statistics – high recovery rates and low mortality rates. Figure 3 shows the word clouds for the first three categories. The distribution of the tweets over the categories identified is shown in Figure 4.

Figure 3. Word Cloud for the Most Common Categories, 1) Constitutional rights and freedom of choice, 2) Conspiracy theory, population control, and big pharma, and 3) Fake news, fake numbers, fake pandemic, and lies.

Figure 4. The Distribution of 51,170 tweets Over Ten Categories Obtained from the LDA Model

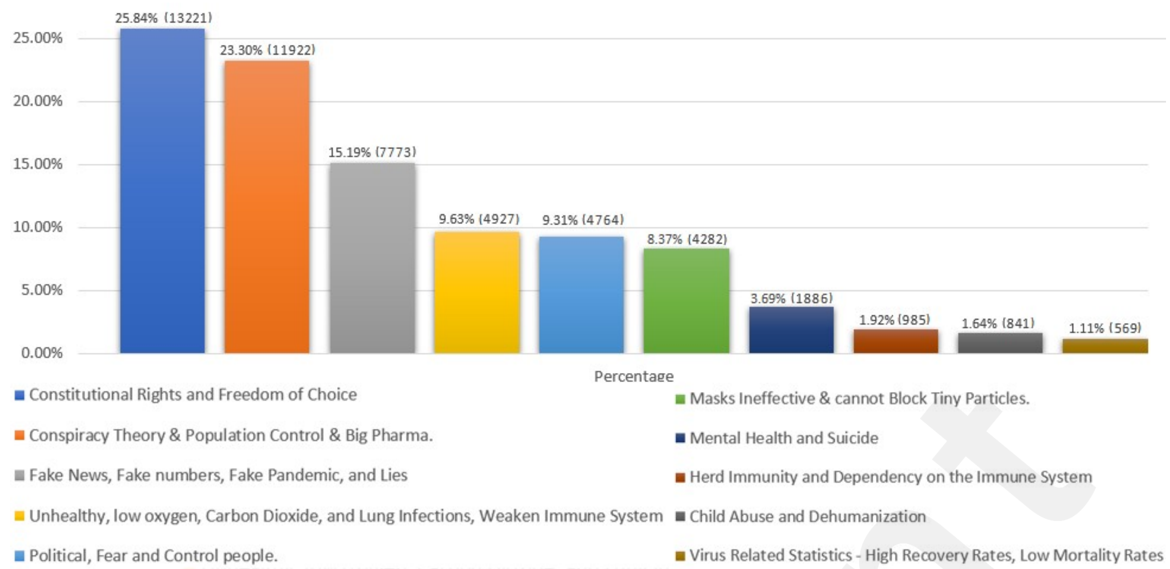
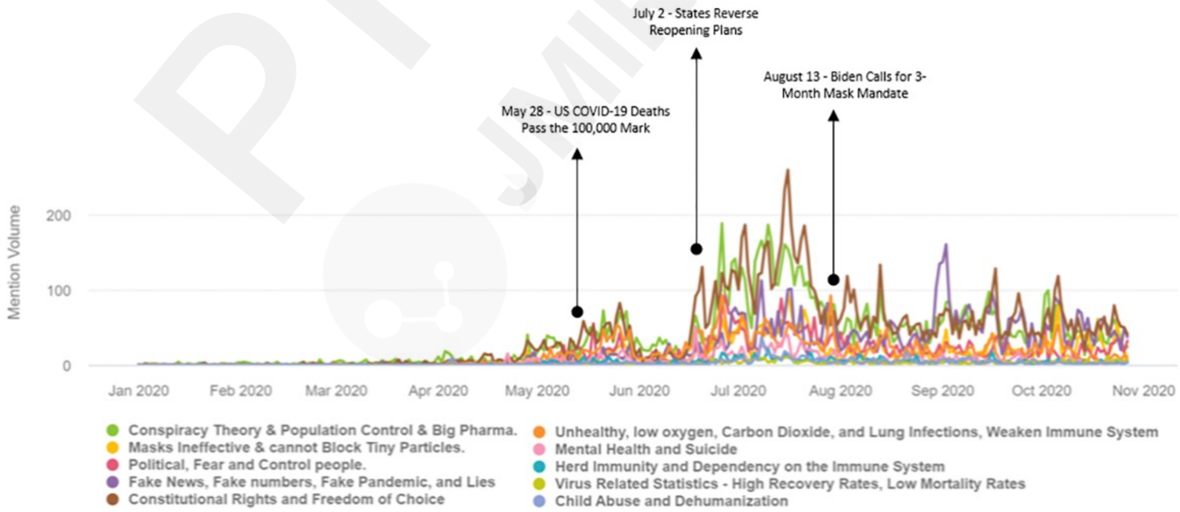


Figure 5 shows the volume of tweets over time by category. Overall, the number of tweets posted increased with time with the highest volume of tweets in July. Between April 8, 2020 and May 29, 2020, a total of 15 states have issued mask mandate [40] which could be related to the spike in tweets posted on masks between April and the beginning of July. Furthermore, between June 18, 2020 and August 11, 2020, 20 more states issued mask mandates [40] and this could explain the increase in tweets posted about masks between late June and mid-august. The figure also shows three relevant milestones between May and August [41]. The three milestones are related to the number of deaths late may, states reversing reopening plans, and the call for 3-month mask mandates. These milestones could also relate to the increasing number of posts on masks. Furthermore, after August 13, we can see consistent debates on masks across all categories.

Figure 5. Tweets Volume and Trend Analysis Over Ten Categories from the LDA Model as well as Three Significant Milestones During the Pandemic Between January 1st, 2020 and October 27th, 2020.



We can also see from figure 5 that more tweets were posted as governments and public health officials requesting people wearing masks as lockdown restrictions were relaxed. The number of tweets posted about constitutional rights and freedom of choice increased noticeably, followed by tweets about conspiracy theory, and population control and big pharma. The following paragraphs provide a synopsis of each of the categories.

Constitutional rights and freedom of choice: Results revealed several reasons that make some Americans refuse to wear face masks despite overwhelming evidence that it saves lives. One important reason discussed during the period of the study is the "Constitutional rights and freedom of choice". Many say mandatory masks violate their constitutional right and freedom of choice. Example tweet: "*Dear #****, I am an American citizen with constitutional rights. I have the right & freedom to choose #NoMask. If u try to enforce this ridiculous order, I will sue your ass 2 hell & back. Kentucky is a #redstate & you don't belong. GTFO. Signed a pissed of Kentucky girl*".
Conspiracy theory, population control and big pharma: Americans also discussed concerns related to conspiracy theory, population control and big pharma. They believed that COVID-19 was human-engineered. Example tweets: "*Won't have to listen to people blabbering on about their latest favourite conspiracy theory,*" "*You can have a ridiculous opinion. Democrats follow blindly, I do not. **** IS Big Pharma. Masks = Control = Submission that will lead to mandatory inoculation of a genetically modifying vaccine. If dems win, we all lose. #MasksOffAmerica*".
Fake news, fake numbers, fake pandemic, and lies: Many also believed the pandemic is fake and there was fake news, misinformation, and lies spread about COVID-19. Example tweets: *"@**** Seasonal flu kills more people EVERY year. You and the fake news media are losing credibility FAST. #nomasks #nonewnormal," "@**** So how many other false positives are out there...this makes the numbers even more questionable*".
Unhealthy, low oxygen, carbon dioxide, and lung infections, weaken immune system: Tweets posted also discussed the health impact of wearing masks. Many believed masks limit oxygen intake and cause rebreathing of carbon dioxide which can lead to lung disease and weaken the immune system. Example tweets: "*Wearing it blocks oxygen and recycles carbon dioxide and carries the bacteria to your respiratory system. #nomasks,*" "*Masks weaken the immune system. Masks allow oral bacteria to affect gums, throat & lungs. Masks limit oxygen intake. Masks cause rebreathing of carbon dioxide,*" "*#COVID-19 #NoMasks Hypercapnia is generally caused by hypoventilation, lung disease, or diminished consciousness*".
Political, fear, and control people: Another topic discussed by Americans was fearmongering. Many believed that politicians and media have only focused on the numbers that gave a negative picture of the pandemic rather than a more balanced and honest look at the numbers. Example tweets: *"@**** Nor do they speak about the low death rate. They want us living in fear. Fear controls the masses! #SheepNoMore #MaskOff," "FEAR MONGERING!!! THIS IS WHAT IT LEADS TO! ENOUGH! NO MORE MASKS!!"*.
Masks ineffective and cannot block tiny particles: Many also had an opinion that masks are ineffective and cannot block tiny particles. Example tweets: "*People wearing #masks and shaming others for NOT wearing them though all #science deems them almost totally ineffective in protecting against the nano particles of the coronavirus. #DumbPandemicDecisions #Masks4All #MasksOff,*" "*@**** says masks are ineffective to stop the virus. Why is there a state execution/executive order now to mandate masks? #NoMasks #ControlRemedy*".
Mental health and suicide: Wearing mask could also have impact on the mental health and could lead to suicidal thoughts. Example tweets: "*they are causing a severe mental health issue. #NoMasks #MasksOff," "Masks are causing horrible harm with the mental health of children. Stop wearing them before these damages are irreversible! #NoMasks #MasksOffArizona," "Masks are causing serious mental health issues in children. Stop with the masks before it's too late! #MasksOff," "Where is the **** physician saying that this lockdown needs to end b/c suicide is up? Mental health has been ignored completely*".
Herd immunity and dependency on the immune system: People should not be forced to wear masks in order to build herd immunity and maintain a healthy and strong immune system. Example tweets: "*It's time we focus on REAL solutions like herd immunity. #NOMASK for me. @****,*" "*You need INTERACTION with people and #NoMasks to maintain a healthy immune*

system #OpenAmericaNOW #OPenHawaiiNow,” “I will NOT wear a damn mask!! It is my right to come in to contact with germs that strengthen my immune system!”.
Child abuse and dehumanization: Asking children to wear masks is child abuse according to many. Example tweets: “*Masking children is child abuse! Kids are not at risk and not carriers of the virus! Kids need to see and communicate clearly. They need to see facial expressions. A mask desensitizes kids! #maskingchildrenischildabuse*,” “*Mandating our young children to wear a mask for 7hrs per day while attending school is tantamount to child abuse. #OpenTheSchools #NoMasks,” “Masks in this case are a tool for soft torture and dehumanization #NoMasks*”.
Virus related statistics (*high recovery rates and low mortality rates)*: People also discussed the high recovery and low mortality rate of the virus which make wearing mask not necessary. Example tweets: *“I will not comply and wear a useless mask that has potential health risks to me for a virus that has a 98% recovery rate. #NoMask,” “COVID-19 Mortality Rate in CA is .00006925% that means 99.999932% are forced 2 destroy R lives 4the weakest virus on the planet! Stop Quarentining the Healthy, Open up Businesses & only Quarantine the Sick! #UnMaskAmerica”.*

Tweet Vs. New COVID-19 cases

Figure 6 depicts the volume of tweets against wearing masks and the number of COVID cases over the study period. The two-time series exhibit a high positive Pearson correlation of 0.77. Since Figure 6 does not provide information about directionality between the two waves, leading and following, we further studied this relationship between both waves in Figure 7. Overall, the results show nine days lead for the volume of tweets over the number of new COVID-19 cases. The 9 days lag is considered comparable to the number of days after which people can develop symptoms of COVID-19. According to [42], approximately, 97% of people infected with COVID-19 developed symptoms within 12 days after exposure.

Figure 6. Pearson Correlation of Tweets against Wearing Masks and New COVID-19 Cases Over Days Between January 2020 and October 2020

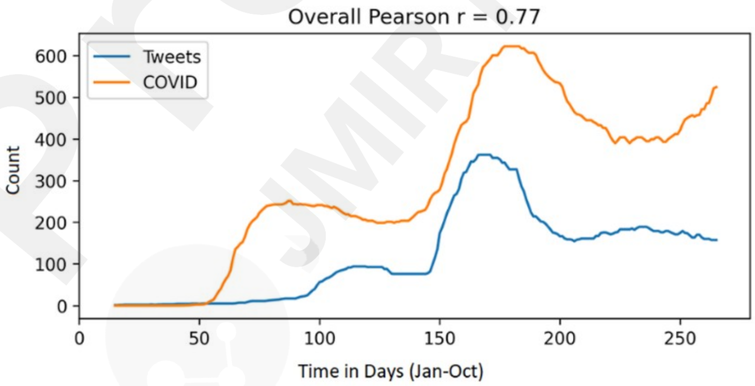
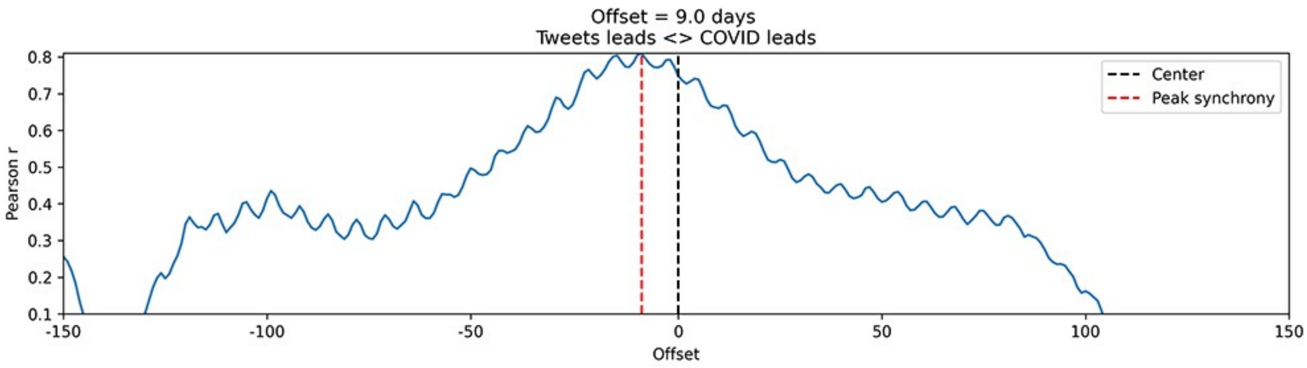


Figure 7. The Volume of Tweets Against Wearing Masks is Leading COVID-19 New Cases by Nine Days



Discussion

This study analyzed the negative stance regarding masks on social media, the specific themes within this discourse, and how this discourse could be associated with the prevalence of new COVID-19 cases. The study reported users' concerns related to constitutional rights and freedom of choice, conspiracy theory, misinformation, health issues, fearmongering and others related to the use of face masks. Further, the time series analysis demonstrated a strong correlation between the number of tweets against wearing masks and the actual number of COVID-19 new cases with the volume of negative tweets is leading the number of new COVID-19 cases by nine days.

The study findings emphasize the potential relationship between social media behavior and the manifestation of this behavior in the physical world. Such finding highlights the importance of listening to social media and proactively reacting to public perception in fighting COVID-19. Lyu and George [43] showed that mask mandates in a number of states were associated with lowering infection rates between '0.9% and 2%' after '1 to 21 days' after wearing masks. However, when the government mandates mask wearing in public, many people feel their constitutional rights and freedom of choice are being violated [44]. As a result, there is a need to increase awareness about the fact that wearing masks can protect others from contracting COVID-19 even though they do not fully protect the person wearing from the infection [45]. On the other hand, the government should also address the challenges faced by implementing a balanced mask wearing mandate that takes into consideration protecting people's lives and protecting their freedom of choice at the same time [46].

Social media platforms have been used to spread fake news, lies, and conspiracy theories, which has a strong impact on people and society [47]. Such impact can cause the public to less likely view actions like wearing masks as a necessity to mitigate the spread of the virus during a pandemic [48]. Therefore, it is crucial that, as we seek to control the spread of COVID-19 and future viruses, we come up with policies to fight against misleading and damaging conspiracy rhetoric. Similarly, there should be policies in place to combat fake news, lies, and misinformation, especially on social media, that could negatively affect trust in science [48].

Healthcare professionals should actively engage in the conversation with the public in order to discuss scientific evidence supporting the importance of wearing a mask and debunk rumors on social media that promotes discussions related to masks causing low oxygen levels or lung infections. Healthcare professionals should discuss evidences and guidelines such as "Wearing a mask does not raise the carbon dioxide (CO_2) level in the air you breathe" [49] and "people age 2 and older should wear masks in public settings and when around people who don't live in their household" [49] to increase awareness regarding the effectiveness of masks in protecting the wearer from inhaling and spreading airborne particles.

Specific children's age groups should be encouraged to wear masks to protect them from COVID-19. However, protecting such age groups by using a mask could be very difficult [50]. To overcome these difficulties, there is a need to advocate for parental involvement and support for the initiatives aimed at increasing mask wearing among children [50]. Children should be encouraged to "take off their masks to breathe in fresh air after wearing masks for a certain amount of time" and should not wear masks in certain cases such as exercising [50]. In the case of non-compliance, it would be a better option not to wear masks by children and follow other measures to reduce infection risk and remain at home [50].

Fearmongering where some people believe that media as well as some governments have focused on numbers that give a negative picture of the pandemic rather than a more balanced and honest view of the numbers [51]. Instead, following an empathetic approach to motivate people to wear masks and adhere to physical distancing could be an effective alternative [52]. In addition, policy makers could use positive messages to curb the spread of fear while still maintaining a transparent and accurate depiction of the situation [51].

With physical, mental, social, and economic burdens imposed by the pandemic, many populations may experience increased suicide risk [53]. Furthermore, high rates of symptoms of anxiety, depression, post-traumatic stress disorder, and stress were reported in a number of countries during the COVID-19 pandemic [54]. Data analysis and events surveillance during the first 6 months of the pandemics have shown specific effects on suicide risk [53]. As a result, knowing the facts about masks and stopping the spread of rumors can reduce stress and adverse impact on mental health [55]. Last but not least, as many believe that herd immunity is the best solution to the crisis and to strengthen their immune systems, a scientific and facts driven view should be shared with public explaining why herd immunity is not an ideal solution as many researchers reported [56].

By carefully analyzing social media posts, policy and decision makers are in a better position to tailor the public health awareness campaigns respond to specific themes and thereby improve its effectiveness in a crisis situation such as that of the COVID-19 pandemic. Along these lines, exploring the categories of tweets surrounding the mask-wearing during the COVID-19 pandemic help reveal a number of insights that could help better design and implement awareness campaigns.

Limitation and Future Work

This study is not without limitations that could be addressed in future works. First, although we identified a very strong correlation between the increase in the volume of tweets against wearing masks and the rise in COVID-19 cases, a limitation remains is that we cannot claim causality, as the cause of the rise in COVID-19 cases could be attributed to other factors such as population density, government lock down restrictions, and others which are out of the scope of this study. Second, the study focused on analyzing English tweets in the United States. Future studies need to address and compare the public discourse on masks across different social media platforms and in different countries. Third, given the number of collected tweets and the focus on Twitter as a data source, the public discourse might not reflect actual public opinion against masks. According to Wojcik and Hughes [57], Twitter has been found to contain much younger audiences with the most prolific 10% of users creating 80% of tweets. Finally, the study did not analyze opinions against mask in early and later stages of the pandemic separately. Such analysis could unmask other important trends that are not present in the paper.

Conclusion

In this study, we analyzed tweets against wearing masks on social media to understand topics, insights, and information about user reported issues. Using data analytics, this study identified trending themes and topics of concern by the public about wearing a mask. The most discussed issues were related to constitutional rights and freedom of choice, conspiracy theory, misinformation, health issues, fearmongering, and ineffectiveness of masks, followed by issues related to mental health, herd immunity, child abuse, and virus-related statistics. Another key finding of this research is highlighting the strong correlation between the increase in the volume of tweets against wearing masks and new COVID-19 cases and the lead of negative tweets before the rise in new COVID-19 cases in the time series analysis. In effect, the findings demonstrated the impact of social media not only on people's opinion or perceptions about public topics, but also the potential impact on real events like changing the course of the pandemic. The significance and implication of this research transcend the COVID-19 pandemic as it demonstrates the importance of social media mining and its potential to support public health-related policies and decisions. Government officials and decision makers could tailor and fine tune the public awareness campaign and prioritize the policy intervention toward the most discussed topics. In case of future massive-scale threats such as the COVID-19 pandemic, government officials and policy makers could leverage social media analytics and surveillance as important tools in proactively responding to the impending crisis. Policy makers need to proactively address public perception and work on shaping this perception through raising awareness, debunking negative sentiments, and adopting early policy intervention to steer the wheel towards public acceptance of more precautionary measures and hence, containing the situation.

Conflicts of interests

The authors do not have any conflict of interest.

References

1. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. The Lancet Respiratory Medicine. 2020;8(5):434-6.
2. Hannah Ritchie EO-O, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, and Max Roser. Mortality Risk of COVID-19. 2021 [cited 2021 Jan 22]; Available from: https://ourworldindata.org/mortality-risk-covid.
3. Fisman DN, Greer AL, Tuite AR. Bidirectional impact of imperfect mask use on reproduction number of COVID-19: A next generation matrix approach. Infectious Disease Modelling. 2020;5:405-8.
4. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. Bmj. 2020;369.
5. MacIntyre CR, Chughtai AA. A rapid systematic review of the efficacy of face masks and respirators against coronaviruses and other respiratory transmissible viruses for the community, healthcare workers and sick patients. International Journal of Nursing Studies. 2020:103629.
6. Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen H-M, et al. Face masks against COVID-19: an evidence review. 2020.
7. Li T, Liu Y, Li M, Qian X, Dai SY. Mask or no mask for COVID-19: A public health and market study. PloS one. 2020;15(8):e0237691.
8. Zhou ZG, Yue DS, Mu CL, Zhang L. Mask is the possible key for self-isolation in COVID-19

pandemic. Journal of Medical Virology. 2020.
9. Ma QX, Shan H, Zhang HL, Li GM, Yang RM, Chen JM. Potential utilities of mask‐wearing and instant hand hygiene for fighting SARS‐CoV‐2. Journal of medical virology. 2020.
10. Wang J, Pan L, Tang S, Ji JS, Shi X. Mask use during COVID-19: A risk adjusted strategy. Environmental Pollution. 2020:115099.
11. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. JMIR Public Health and Surveillance. 2020;6(2):e19273.
12. Zhang Y, Fan Y, Ye Y, Li X, Zheng W, editors. Detecting opioid users from twitter and understanding their perceptions toward mat. 2017 IEEE International Conference on Data Mining Workshops (ICDMW); 2017: IEEE.
13. Wang F-Y, Carley KM, Zeng D, Mao W. Social computing: From social informatics to social intelligence. IEEE Intelligent systems. 2007;22(2):79-83.
14. Wahbeh A, Nasralah T, El-Gayar O, Al-Ramahi M, Elnoshokaty A, editors. Adverse Health Effects of Kratom: An Analysis of Social Media Data. Proceedings of the 54th Hawaii International Conference on System Sciences; 2021.
15. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: mixed methods analysis. JMIR public health and surveillance. 2020;6(2):e19276.
16. El-Gayar O, Nasralah T, Elnoshokaty A, editors. Wearable devices for health and wellbeing: Design Insights from Twitter. Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS-52); 2019; Hawaii, USA.
17. Runge KK, Yeo SK, Cacciatore M, Scheufele DA, Brossard D, Xenos M, et al. Tweeting nano: How public discourses about nanotechnology develop in social media environments. Journal of nanoparticle research. 2013;15(1):1381.
18. Nzali MDT, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR medical informatics. 2017;5(3):e23.
19. Glowacki EM, Lazard AJ, Wilcox GB. E-cigarette topics shared by medical professionals: a comparison of tweets from the United States and United Kingdom. Cyberpsychology, Behavior, and Social Networking. 2017;20(2):133-7.
20. Nasralah T, El-Gayar O, Wang Y. What Social Media Can Tell Us About Opioid Addicts: Twitter Data Case Analysis. 2019.
21. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Grucza R, Bierut L. A content analysis of tweets about high-potency marijuana. Drug and alcohol dependence. 2016;166:100-8.
22. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": Analyzing Twitter data on marijuana concentrates across the US. Drug and alcohol dependence. 2015;155:307-11.
23. Dai H, Hao J. Mining social media data on marijuana use for post traumatic stress disorder. Computers in Human Behavior. 2017;70:282-90.
24. Wahbeh A, Al-Ramahi M, El-Gayar O, Nasralah T. Health Risks of e-cigarettes: Analysis of Twitter Data Using Topic Mining. 2020.
25. Raymond JR. The Great Mask Debate: A Debate That Shouldn't Be a Debate at All. WMJ. 2020;119(4):229-39.
26. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research. 2003;3(Jan):993-1022.
27. Hopkins DJ, King G. A Method of Automated Nonparametric Content Analysis for Social Science. American Journal of Political Science. 2010 2010;54(1):229-47. doi: 10.1111/j.1540-

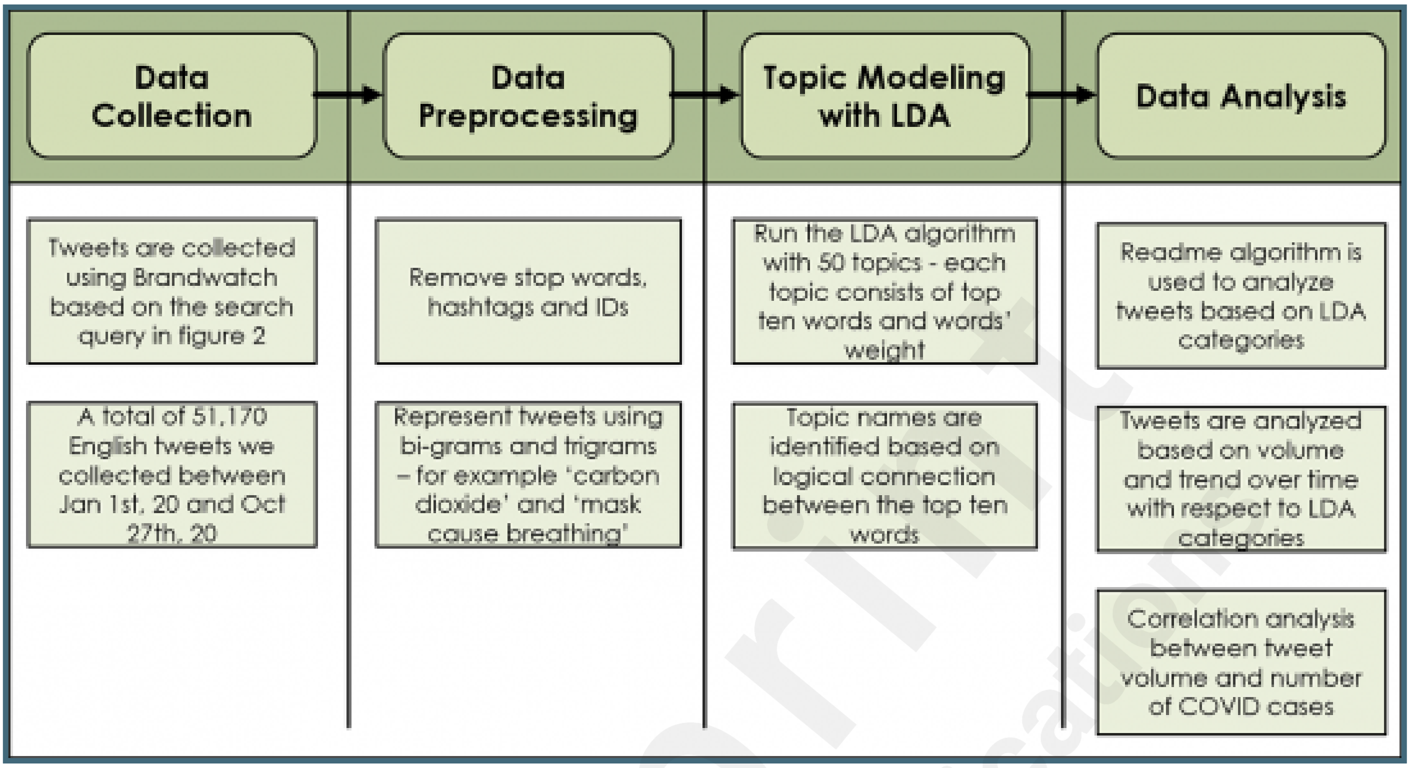
5907.2009.00428.x.
28. Lewiston-Aurburn. Maine's Twitter users among top for anti-mask feelings. sunjournal.com: sunjournal.com; 2020 [cited 2020 Nov 1st]; Available from: https://www.sunjournal.com/2020/07/10/maines-twitter-users-among-top-for-anti-mask-feelings/.
29. Hopkins J. Covid-19 Data in Motion. Johns Hopkins University: Johns Hopkins University; 2020 [cited 2020]; Available from: https://coronavirus.jhu.edu/.
30. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychology, behavior, and social networking. 2013;16(9):708-13.
31. Krotov V, Silva L, editors. Legality and ethics of web scraping. Twenty-fourth Americas Conference on Information Systems (AMCIS); 2018; New Orleans. https://aisel.aisnet.org/amcis2018/DataScience/Presentations/17: AIS.
32. Le Q, Mikolov T, editors. Distributed representations of sentences and documents. International conference on machine learning; 2014.
33. Wahbeh A, Al-Ramahi M, Noteboom C, Nasralah T, editors. Discovering Patient Portal Features Critical to User Satisfaction: A Systematic Analysis. Proceedings of the 52nd Hawaii International Conference on System Sciences; 2019.
34. Al-Ramahi MA, Liu J, El-Gayar OF. Discovering design principles for health behavioral change support systems: A text mining approach. ACM Transactions on Management Information Systems (TMIS). 2017;8(2-3):1-24.
35. Al-Ramahi M, Noteboom C. Mining User-generated Content of Mobile Patient Portal: Dimensions of User Experience. ACM Transactions on Social Computing. 2020;3(3):1-24.
36. Noteboom C, Al-Ramahi M. What are the Gaps in Mobile Patient Portal? Mining Users Feedback Using Topic Modeling. 2018.
37. Al-Ramahi M, Noteboom C, editors. A Systematic Analysis of Patient Portals Adoption, Acceptance and Usage: The Trajectory for Triple Aim? Proceedings of the 51st Hawaii International Conference on System Sciences; 2018. https://scholarspace.manoa.hawaii.edu/handle/10125/49994.
38. Mudelsee M. Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. Mathematical Geology. 2003;35(6):651-65.
39. Haugh LD. Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. Journal of the American Statistical Association. 1976;71(354):378-85.
40. Kim A. These are the states requiring people to wear masks when out in public. CNN; 2020 [cited 2021 Jan 20, 2021]; Available from: https://www.cnn.com/2020/06/19/us/states-face-mask-coronavirus-trnd/index.html.
41. AJMC. A Timeline of COVID-19 Developments in 2020. AJMC; 2021 [cited 2021 Jan 31]; Available from: https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020.
42. Tavernise S. It may be weeks before we know if Thanksgiving travel fed a virus surge. The new York Times; 2020; Available from: https://www.nytimes.com/2020/11/30/world/thanksgiving-coronavirus-us-spread.html.
43. Lyu W, Wehby GL. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US: Study examines impact on COVID-19 growth rates associated with state government mandates requiring face mask use in public. Health affairs. 2020;39(8):1419-25.
44. Scerri M, Grech V. To wear or not to wear? Adherence to face mask use during the COVID-19 and Spanish influenza pandemics. Early human development. 2020:105253.

45. Betsch C, Korn L, Sprengholz P, Felgendreff L, Eitze S, Schmid P, et al. Social and behavioral consequences of mask policies during the COVID-19 pandemic. Proceedings of the National Academy of Sciences. 2020;117(36):21851-3.
46. Lenhart B. The Peoples' Constitution: COVID-19 versus Freedom. https://loudounnow.com/: https://loudounnow.com/; 2020 [cited 2020 Dec 2nd]; Available from: https://loudounnow.com/2020/05/07/the-peoples-constitution-covid-19-versus-freedom/.
47. Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter. 2017;19(1):22-36.
48. Palm R. Conspiracy Theories About The Origins Of COVID-19 Outweigh Science's Influence, Researchers Say. https://news.gsu.edu/: Geogia State University news; 2020 [cited 2020 Dec 2nd]; Available from: https://news.gsu.edu/2020/10/29/covid-19-conspiracy-theory-exposure-risa-palm/.
49. CDC. Considerations for Wearing Masks Help Slow the Spread of COVID-19. cdc.gov: cdc.gove; 2020 [cited 2020]; Available from: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html.
50. Esposito S, Principi N. To mask or not to mask children to overcome COVID-19. European journal of pediatrics. 2020;179(8):1267-70.
51. Moreno FsR. COVID-19: The Fearmongering Must End. frcaction.org: frcaction; 2020 [cited 2020 Dec 2nd]; Available from: https://www.frcaction.org/updatearticle/20201028/covid-fearmongering.
52. Pfattheicher S, Nockur L, Böhm R, Sassenrath C, Petersen MB. The emotional path to action: Empathy promotes physical distancing and wearing of face masks during the COVID-19 pandemic. Psychological Science. 2020;31(11):1363-73.
53. Moutier C. Suicide prevention in the COVID-19 era: Transforming threat into opportunity. JAMA psychiatry. 2020.
54. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. Journal of affective disorders. 2020.
55. CDC. Coping with Stress. cdc.gov: cdc; 2020 [cited 2020 Dec 2nd]; Available from: https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html.
56. Aschwanden C. The false promise of herd immunity for COVID-19. https://www.nature.com/: https://www.nature.com/; 2020 [cited 2020 Dec 22nd]; Available from: https://www.nature.com/articles/d41586-020-02948-4.
57. Wojcik S, Hughes A. Sizing up Twitter users. PEW research center. 2019;24.

Supplementary Files

Figures

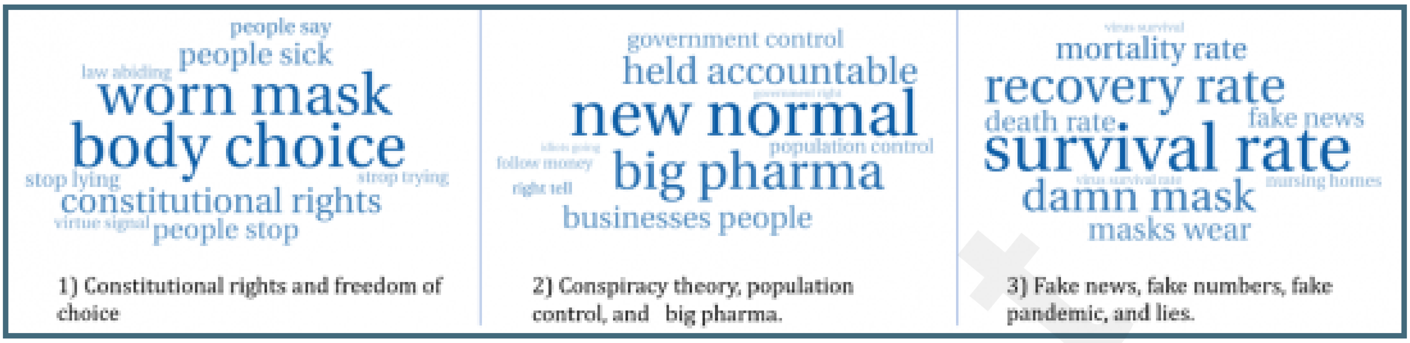
Methodology for Mining Social Media.



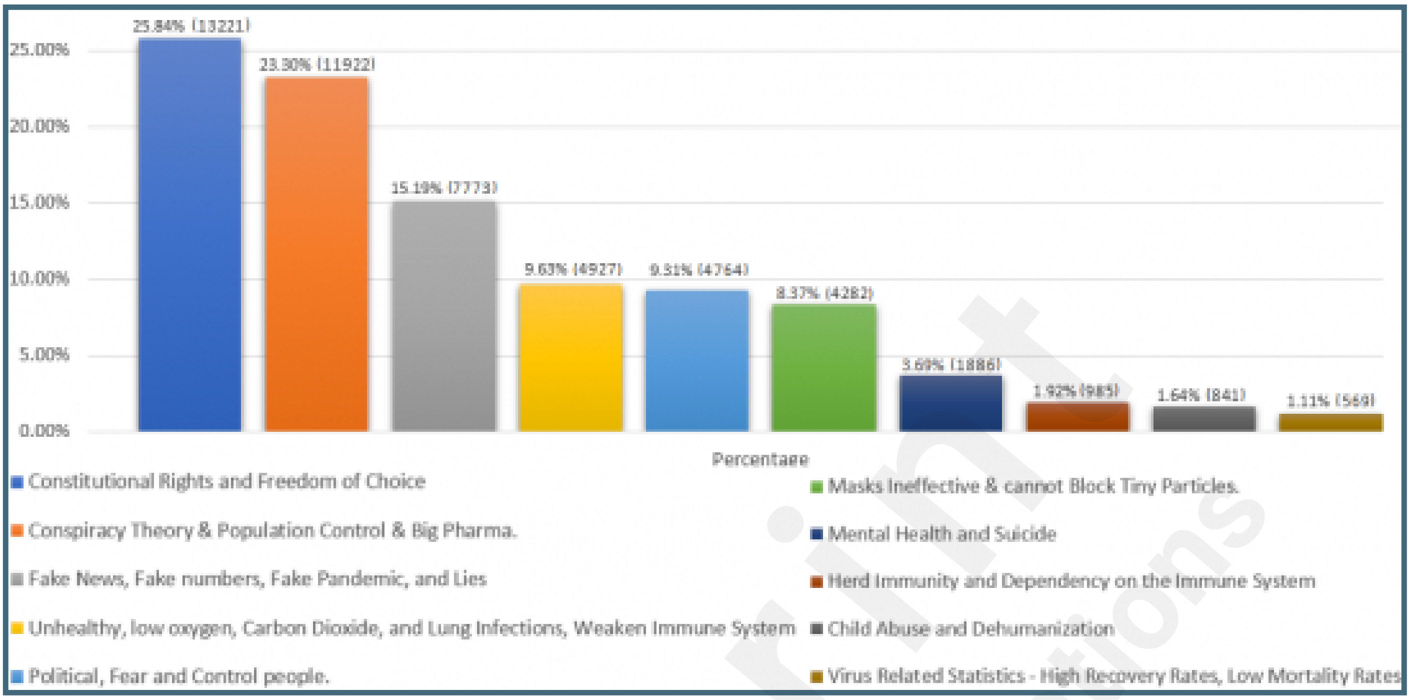
Hashtags and Search Query Used for Data Collection.

```
(#maskoffamerica  OR #NoMask* OR #takeoffthemask OR #maskburning OR #BurnYourMask OR
#burnyourmaskchallenge OR #masksdontwork OR #nomaskonme OR #NoMasksOnMe OR #nomaskselfie OR
#maskhoax OR #maskshoax OR #NoMasksEVER  OR #NoMoreMasks OR #IWillNotWearAMask OR #MaskOff*
OR #MasksOff* OR #facefreedom  OR #masksmakemesweaty OR #MasksAreDangerous OR #TakeTheMaskOff
OR #Stopforcingmasksonme OR #takeoffyourmask* OR #refusemask* OR  #NeverMasker OR
#StopWearingMask* OR #StopWearingTheDamnMasks OR #MasksdontMatter OR #stopmasking OR
#stopthestupidmask OR #maskingchildrenischildabuse OR #MomsAgainstMasks OR #MasksRUnhealthy OR
#SheepWearMasks OR #MasksAreForSheep OR  #RefuseToWearMasks OR #MasksAreMurderingMe)
AND - (RT OR http OR https)
```

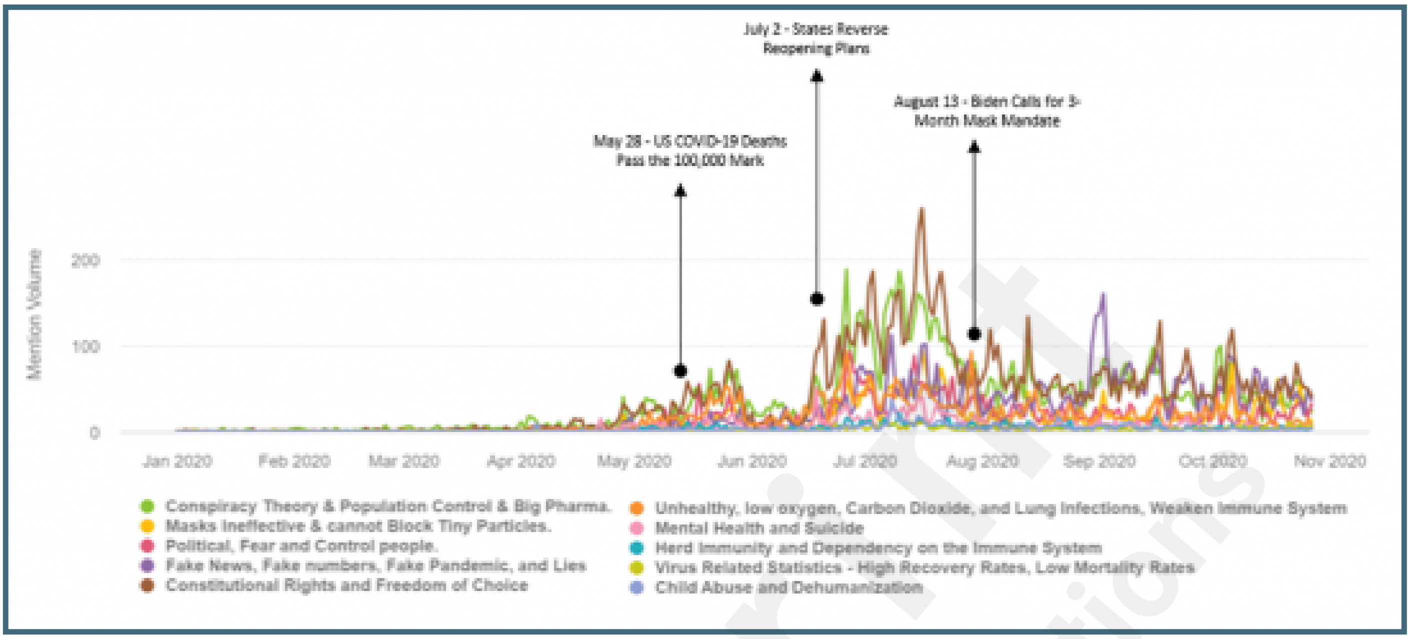
Word cloud for the most common categories, 1) constitutional rights and freedom of choice, 2) conspiracy theory, population control, and big pharma, and 3) fake news, fake numbers, fake pandemic, and lies.



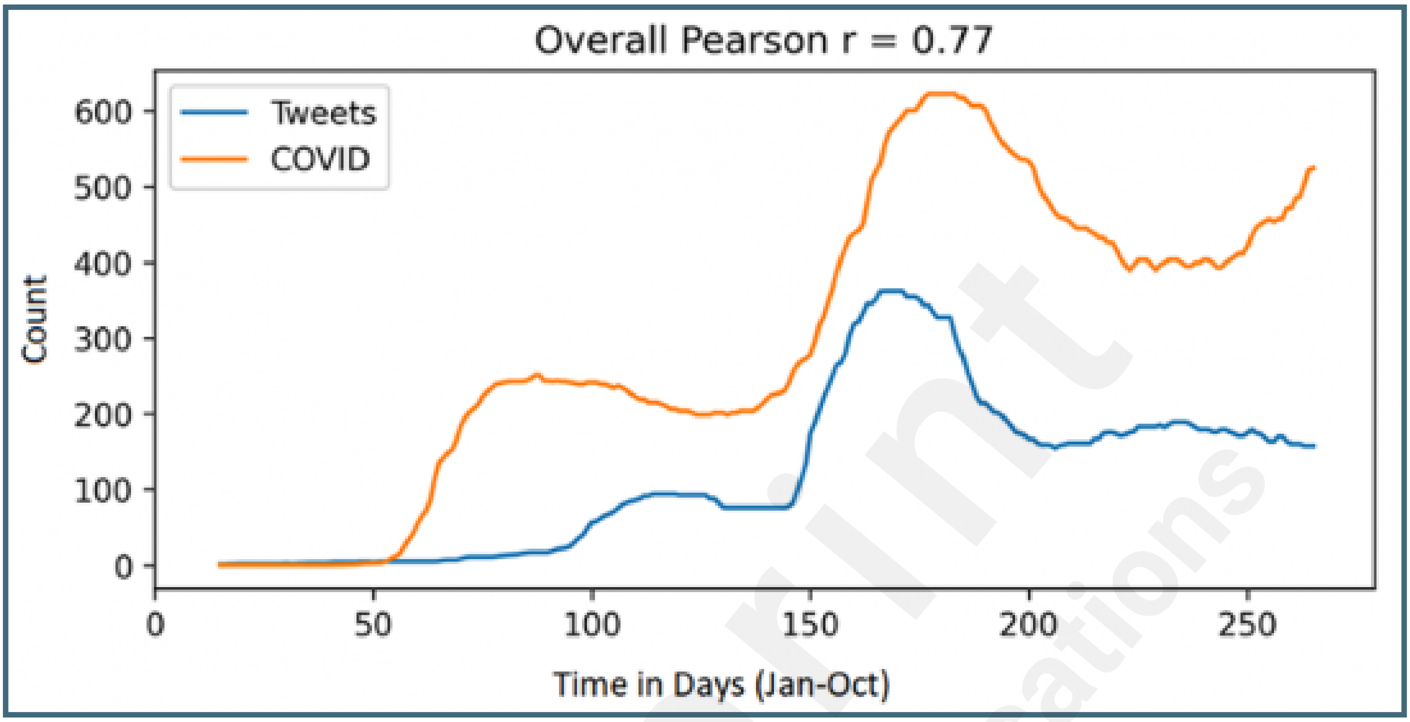
The Distribution of 51,170 tweets Over Ten Categories Obtained from the LDA Model.



Tweets Volume and Trend Analysis Over Ten Categories from the LDA Model as well as Three Significant Milestones During the Pandemic Between January 1st, 2020 and October 27th, 2020.



Pearson Correlation of Tweets against Wearing Masks and New COVID-19 Cases Over Days Between January 2020 and October 2020.



The Volume of Tweets Against Wearing Masks is Leading COVID-19 New Cases by Nine Days.