Association for Information Systems AIS Electronic Library (AISeL)

AMCIS 2020 Proceedings

Data Science and Analytics for Decision Support (SIGDSA)

Aug 10th, 12:00 AM

Early Public Outlook on the Coronavirus Disease (COVID-19): A Social Media Study

Giridhar Reddy Bojja Dakota State University, giridhar.bojja@trojans.dsu.edu

Martinson Ofori Dakota State University, martinson.ofori@trojans.dsu.edu

Jun Liu Dakota State University, jun.liu@dsu.edu

Loknath Sai Ambati Dakota State University, loknathsai.ambati@trojans.dsu.edu

Follow this and additional works at: https://aisel.aisnet.org/amcis2020

Recommended Citation

Bojja, Giridhar Reddy; Ofori, Martinson; Liu, Jun; and Ambati, Loknath Sai, "Early Public Outlook on the Coronavirus Disease (COVID-19): A Social Media Study" (2020). *AMCIS 2020 Proceedings*. 31. https://aisel.aisnet.org/amcis2020/data_science_analytics_for_decision_support/ data_science_analytics_for_decision_support/31

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Early Public Outlook on the Coronavirus Disease (COVID-19): A Social Media Study

Completed Research

Giridhar Reddy Bojja Dakota State University <u>giridhar.bojja@trojans.dsu.edu</u> Martinson Ofori Dakota State University <u>martinson.ofori@trojans.dsu.edu</u>

Jun Liu Dakota State University jun.liu@dsu.edu Loknath Sai Ambati Dakota State University loknathsai.ambati@trojans.dsu.edu

Abstract

The recent outbreak of the coronavirus (COVID-19) brought with its public concerns and fears about a global epidemic. With the increase in the popularity, usage, and reach of social media, this research examined the early public outlook on COVID-19 using SM-Platform, Twitter.com. The current study employed a mixed-method approach in collecting and analyzing public tweets by combining quantitative sentiment analysis with a qualitative thematic analysis. Our results revealed positive sentiment prior to the spread of the disease. The sentiment then turned negative as the disease spread, accompanied by a large amount of fear as rumors. In a thematic analysis we also uncovered nine key topics on the disease including, but not limited to, prevention, symptoms and spread of disease. Our study will provide an understanding of social media and public health outbreak surveillance. The findings of the research revealed the usefulness of twitter mining to illuminate public health education.

Keywords

CoVID-19, Coronavirus, Social media, Twitter, Sentiment Analysis, Thematic analysis, Health Communication, China.

Background

In the past two decades, two main types of coronaviruses have been discovered globally – Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). Coronaviruses have previously been identified in mice, chicken and other animals and often caused life-threatening pneumonia when identified in humans (van der Hoek et al. 2004). However, at the end of 2019, unknown strains of pneumonia were detected in the Chinese city of Wuhan (WHO 2020a). Subsequently the disease was identified as a novel coronavirus which had not been previously identified. The World Health Organization (WHO) announced an official name for the disease to be Coronavirus Disease 2019 (COVID-19) while also indicating the disease to be caused by a virus named SARS-CoV-2. COVID-19 causes respiratory disease extending from basic cold to extreme intense respiratory disorder (SARS). Common symptoms include fever, cold and difficulty in breathing (CDC 2020). At the time of this report, no known vaccine had been developed as a means of prevention, however, to prevent further spread of the disease, the WHO continues to work closely with global experts, governments and partners to gain more knowledge on the disease (WHO 2020a).

The fast-paced spread of the disease and the morbidity associated with contracting COVID-19 has caused a global panic. In many cases, people turn to social media to post or seek out news on the disease. Among the various social media websites, Twitter.com is widely used for discussions on public health (Neiger et al. 2012). It is viewed as a medium for exchange of news and information regarding public health events, confirmed by its adequacy for analyzing events during the H1N1 pandemic planning activities (Sullivan et

al. 2012). In the time of global emergencies, public respond, react and mobilize on the situation in order to help people in need (Palen et al. 2010). Twitter's ability for wide reach, functionality, and low costs can possibly catch pandemic patterns, gather data, and disperse knowledge. The utilization of twitter underpins its capability to impact public health outbreak surveillance endeavors in a new and inventive manner (Khan et al. 2010). It has also been used in epidemic tracking for general influenza, malaria, and Ebola outbreaks (Boit and El-Gayar 2020; Broniatowski et al. 2015; Deiner et al. 2016; Lazard et al. 2015; Signorini et al. 2011). Given the distress and public outcry associated with the spread of the COVID-19, there is an inherent gap on where the public understanding of the disease lies. On issues of public health, there is often a disparity between credible news shared by stakeholders and the public sentiments. The general varying perspectives makes it difficult to empirically measure public understanding of issues raised during distressing situations such as disease outbreaks. This study aims to use social media analytics to close the inherent gap in extant literature by investigating the notion and general public sentiment on the topic of COVID-19. Accordingly, this study will explore the behavioral panorama of the social media users and imminent discussions around early outlook on CoVID-19. Therefore, we investigate following research questions:

- What is the net sentiment and emotion change over time on COVID-19 on Twitter?
- What are the prevalent themes discovered over time on Twitter regarding COVID-19?

Methods

This study employs a mixed method approach. This approach allows one to identify aspects of a phenomenon more accurately by approaching it from different vantage points using different methods and techniques. A quantitative approach employing machine learning techniques was used to collect the data and perform sentiment and emotion analysis. Qualitative approach was used to perform thematic analysis.

Data Sources and Search Strategy

This study employs social media analytics to identify public sentiments, emotion and related topics of interest on COVID-19. We mined publicly available social media data using Crimson Hexagon's ForSight platform (Crimson Hexagon 2019a). The advantage of using Crimson Hexagon software is that it facilitates data collection and provides insights into online conversations in terms of sentiments and emotions. The ForSight platform has access to the full Twitter Firehose (through a direct partnership with Twitter) and provides access to complete coverage of all tweets globally since 2010. We mined data starting 6 months before the earliest reports of the disease appeared in news and other outlets online. In effect, our search query targeted tweets on COVID-19 posted between July 2019 and February 2020. Our search query targeted various names the COVID-19 has been referred to since its detection. We use the full name (coronavirus), the initial provisional name (2019-nCov), and the new official name of this novel strain of the disease (COVID-19) (CDC 2020). To enable the capture of tweets before various names were given, we also search for tweets about pneumonia that mentions specifically Wuhan or China. We did not add any exclusion terms to the initial search in order to gather as much volume as possible. Figure 1 shows the specific keywords used in our search.

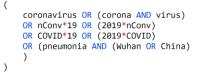


Figure 1. Search query

Data Analysis

The ForSight platform uses Buzz monitors as pretrained classifiers that perform unsupervised clustering, sentiment and emotion analysis. During sentiment analysis, posts are classified as positive, negative or neutral. ForSight performs sentiment analysis using a vast training set of over 500,000 hand-labelled posts (Crimson Hexagon 2019b). Emotion analysis is also conducted on the platform by categorizing posts into six feelings of emotion based on the *Ekman 6: anger, fear, disgust, joy, surprise,* and *sadness* (Crimson Hexagon 2018).

To identify the prevalent themes in the tweets, we use the Cluster Module. This module is one of four visualizations on available on ForSight that samples words from 1,000 up to 10,000 posts per day for the chosen time frame for unsupervised clustering (Crimson Hexagon 2019c). The platform then arranges them into interconnected bubbles as a representation of the relationships between the words in online conversations.

Some news outlets have compiled a comprehensive timeline on the COVID-19 outbreak (Al Jazeera 2020; Antonio 2020; CNN Editorial Research 2020). Using these sources as a guide, we analyze the twitter activity in a systematic manner consisting of four phases:

- **Phase 1 (July 1, 2019 December 11, 2019).** This is the entire phase before the first reports of pneumonia starts to surface in Wuhan inside mainland China.
- **Phase 2 (December 12, 2019 January 4, 2020).** Also known as the rumor phase, this phase analyzes activity on twitter when the virus was still not known. Several differing opinions existed at this time on what the disease could be.
- **Phase 3 (January 5, 2019 January 29,2020).** Here, we look to identify what happens after an official report released by government website in Wuhan reports 59 cases of the disease (7 of them severe) while stating the disease is neither SARS-CoV or MERS-CoV¹.
- **Phase 4 (January 30 February 2020).** In this phase, we examine the effect after the WHO declaration of COVID-19 as a "*public health emergency of international concern*" (WHO 2020b).

Thematic Analysis

Thematic analysis emphasizes identifying, analyzing and interpreting patterns of meaning within the data. Using a randomly selected sample of 5,000 tweets exported from ForSight, we also conduct a thematic analysis of the tweets using the MAXQDA software. Random sampling performed by ForSight guarantees that there is an equivalent possibility of choosing any single tweet. We invited two PhD students that are not related to the research to code all 5,000 tweets. We calculated the inter-evaluator agreement with Cohen's Kappa (Viera et al., 2005). A Kappa value of 1 indicates perfect agreement whereas a Kappa value of 0 or less indicates agreement by chance.

Discussion and Results

COVID-19 outbreak over the period

We mined a total of 40,759,360 posts between 1st July 2019 and 20th February 2020. As shown in Figure 2, on most days prior to Phase 1 activity, there were less than 150 posts, discussing the disease or cases of pneumonia. Phase 2 also shows moderate twitter activities with the highest number of tweets occurring between December 15th.to December 17th. In Phase 3, January 5th sees a moderate rise in number of tweets, which then goes down in the following days but then starts a gradual rise from January 19th. Phase 4 shows heavy twitter activities with the days before February 11th showing over a million tweets a day and about an average of 750,000 tweets a day afterwards



Figure 2. Trend of discussion on COVID-19

¹ <u>http://wjw.wuhan.gov.cn/front/web/showDetail/2020010509020</u>

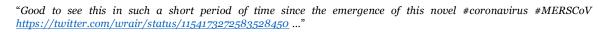
Predictably, about 76% of users expressed negative sentiment on the disease while the primary feeling of emotion was *Disgust* (a very strong feeling of not liking something). Surprisingly however, the second most prevalent emotion was that of *joy*, then *fear*, *sadness*, *anger* and *surprise* in that order. From the analysis of the tweets, we found the joy posts were mostly due to users tweeting encouraging words to healthcare providers and researchers working fervently to contain the disease. Below are some example tweets that show different sentiments.

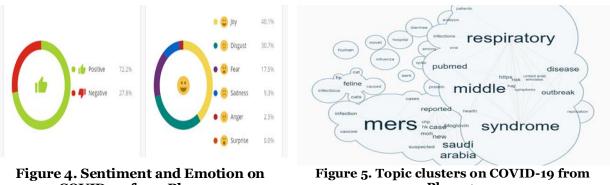
Sentimental Posts

| Schunchta | |
|--------------------|---|
| Negative | "Upsetting that some people are using Coronavirus as an excuse to be racist and discriminatory. It's completely unnecessary and unacceptable! Please stand up for those affected if you witness any racist behaviour #stopracism #coronavirus" |
| Positive | "Amazed at how fast data about this new Wuhan coronavirus is being discovered and shared! Credit to all those public health officials, clinicians and scientists! <u>https://twitter.com/trvrb/status/1216118434850725888</u> " |
| Emotional H | Posts |
| Disgust | "Ewwww! Sick people everywhere! Stay away! O and the #Flu in muriKKKa. #Coronavirus: Second #death from #SARS-like #illness in Wuhan, #China - CNN https://www.cnn.com/2020/01/16/health/coronavirus- wuhan-second-death-intl-hnk/index.html#FridayMorning #sickness" |
| Joy | "It's amazing to see the preparedness response to the #Wuhan coronavirus in Toronto with our partners in public health. Knowing that we are ready to fight is the best feeling, grateful to live in a city that prioritizes the safety and health of populations" |
| Fear | "im so scared of getting sick and believe it or not, the coronavirus shit happening rn is making my health anxiety go MAD. i know the chances of me getting it are slim, but im so scared of being externely sick" |

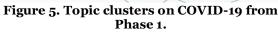
Phase 1 (July 1, 2019 – December 11, 2019)

In this phase, the sentiment of the tweets is 72.2% positive and 27.8% negative. The predominant emotion is joy, followed by disgust and fear. Figure 4 shows the sentiment and emotion within Phase 1. Most positive sentiment in this period came from responses to research efforts towards MERS-CoV vaccines. An example of such tweet is:





COVID-19 from Phase 1.



The topic clusters highlight the MERS (and SARS) strain of the virus. The Middle Eastern origins of the MERS-CoV is emphasized here as we observe that several tweets mention Saudi Arabia and the Hajj (an annual Islamic pilgrimage to Mecca in Saudi Arabia). We also detect a cluster of tweets on cats and felines (reported as the origins of the SARS virus). Figure 5 highlights the topic clusters from this period.

Phase 2 (December 12, 2019 – January 4, 2020)

The total number of posts from this phase of the outbreak were 21,861. There are very few tweets mentioning the disease with less than 200 tweets daily before December 30, 2019. The number of tweets then guickly rose from 255 posts on December 30 to 2,403 posts on December 31, and the number kept rising till the January 4 cutoff point. The sentiments in this period quickly became highly negative with

over 87% of negative posts as compared to 13% of positive posts. With so many rumors flying around the new virus, there is a high amount of fear (61.4%) and disgust (24.8%). The emotion and sentiments for this period are shown in Figure 6. Table 1 compares the sentiments and emotion across the different time-periods.

As we show in Figure 7 the two themes can be inferred from the clusters formed from posts in this phase: 1) **Locations of Outbreaks**. Geographic locations such as Wuhan and China are mentioned several times in tweets on the disease while moderate mentions of Hong Kong and Taiwan are also mentioned. 2)**Unknown Disease.** Mentions of atypical pneumonia, viral pneumonia, new cases, and calls for containments can also be seen in the clusters.





Figure 6. Sentiment and Emotion on COVID-19 from Phase 2.

Figure 7. Topic clusters on COVID-19 from Phase 2.

| | Sentiment | | | | | | | |
|--------------------------------|---|---|--|--|--|---|--|--|
| Main Press Highlight | Negative | Positive | Fear | Disgust | Joy | Surprise | Sadness | Anger |
| Chinese state broadcast CCTV | | | | | | | | |
| reports new viral outbreak was | | | | | | | | |
| in Wuhan, China. | 85.4 | 14.6 | 25.9 | 39.3 | 27.7 | 0.9 | 4.5 | 1.8 |
| China alerts WHO on several | | | | | | | | |
| cases of unusual pneumonia in | | | | | | | | |
| Wuhan. | 93.8 | 6.2 | 68.1 | 18.0 | 11.2 | 0.5 | 0.5 | 1.6 |
| Huanan Seafood Wholesale | | | | | | | | |
| Market closed as possible | | | | | | | | |
| source of virus. | 85.3 | 14.7 | 61.9 | 25.1 | 10.0 | 0.1 | 2.0 | 0.9 |
| | Chinese state broadcast CCTV reports new viral outbreak was in Wuhan, China. China alerts WHO on several cases of unusual pneumonia in Wuhan. Huanan Seafood Wholesale Market closed as possible | Main Press HighlightNegativeChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.4China alerts WHO on several cases of unusual pneumonia in Wuhan.93.8Huanan Seafood Wholesale Market closed as possible93.8 | Main Press HighlightNegativePositiveChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.6China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.2Huanan Seafood Wholesale Market closed as possible14.614.6 | Main Press HighlightNegativePositiveFearChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.9China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.1Huanan Seafood Wholesale Market closed as possible | Main Press HighlightNegativePositiveFearDisgustChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.3China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.0Huanan Seafood Wholesale Market closed as possible </td <td>Main Press HighlightNegativePositiveFearDisgustJoyChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.7China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.2Huanan Seafood Wholesale Market closed as possibleImage: Constant of the second secon</td> <td>Main Press HighlightNegativePositiveFearDisgustJoySurpriseChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.70.9China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.20.5Huanan Seafood Wholesale Market closed as possible<!--</td--><td>Main Press HighlightNegativePositiveFearDisgustJoySurpriseSadnessChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.70.94.5China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.20.50.5Huanan Seafood Wholesale Market closed as possible<!--</td--></td></td> | Main Press HighlightNegativePositiveFearDisgustJoyChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.7China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.2Huanan Seafood Wholesale Market closed as possibleImage: Constant of the second secon | Main Press HighlightNegativePositiveFearDisgustJoySurpriseChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.70.9China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.20.5Huanan Seafood Wholesale Market closed as possible </td <td>Main Press HighlightNegativePositiveFearDisgustJoySurpriseSadnessChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.70.94.5China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.20.50.5Huanan Seafood Wholesale Market closed as possible<!--</td--></td> | Main Press HighlightNegativePositiveFearDisgustJoySurpriseSadnessChinese state broadcast CCTV reports new viral outbreak was in Wuhan, China.85.414.625.939.327.70.94.5China alerts WHO on several cases of unusual pneumonia in Wuhan.93.86.268.118.011.20.50.5Huanan Seafood Wholesale Market closed as possible </td |

Table 1. Highlights from Phase 2.

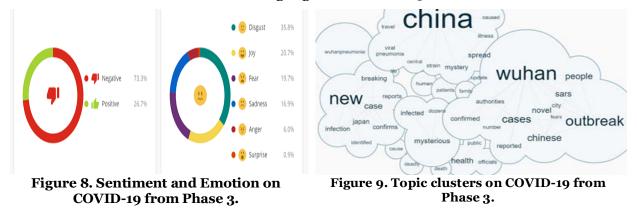
Phase 3 (January 5, 2019 – January 29,2020)

Although this period culminated in over 14M tweets, the ruling-out of the disease of SARS-CoV and MERS-CoV did not spark an immediate increase in number of tweets. Generally, the volume of tweets remained around 5,000 daily up to January 8. The volume of tweets then rose to above 10,000 daily, starting a steep rise on January 18 which peaked at 2.5M tweets on January 26. The sentiments for this period are 73.3% negative and 26.7% positive. As displayed in Figure 8, we notice an emotion change from fear in the previous phase to *disgust* (35.8%), *joy* (20.7%) and *fear* (19.7%).

| | | Sentiment | | | Emotion | | | | | |
|--------------|-------------------------------------|-----------|----------|------|---------|------|----------|---------|-------|--|
| Date | Main Press Highlight | Negative | Positive | Fear | Disgust | Joy | Surprise | Sadness | Anger | |
| January 5 - | Experts rule out the possibility of | | | | | | | | | |
| January 6 | SARS-CoV and MERS-CoV. | 74.8 | 25.2 | 34.1 | 49.9 | 11.1 | 0.5 | 3.5 | 0.9 | |
| | Chinese authorities identify the | | | | | | | | | |
| January 7 - | virus as a novel coronavirus (2019- | | | | | | | | | |
| January 10 | nCoV). | 84.5 | 15.5 | 33.1 | 39.0 | 18.5 | 0.1 | 6.7 | 2.5 | |
| | 61-year-old man dies after | | | | | | | | | |
| January 11 - | respiratory failure caused by | | | | | | | | | |
| January 12 | severe pneumonia. | 74.0 | 26.0 | 26.3 | 16.3 | 31.6 | 0.1 | 25.4 | 0.4 | |

| | First case in outside China | | | | | Ì | | | |
|--------------|-------------------------------------|------|------|------|------|------|-----|------|-----|
| January 13 - | reported in Thailand from a | | | | | | | | |
| January 15 | woman arriving from Wuhan. | 81.7 | 18.3 | 31.6 | 33.4 | 28.7 | 0.1 | 4.6 | 1.6 |
| | First case inside Japan from a man | | | | | | | | |
| January 16. | also arriving from Wuhan. | 95.1 | 4.9 | 35.7 | 32.5 | 15.5 | 0.1 | 15.1 | 1.1 |
| | Second death reported in China. | | | | | | | | |
| January 17 - | Three US airports start symptoms | | | | | | | | |
| January 19 | screening. | 94.1 | 5.9 | 26.6 | 44.4 | 16.1 | 0.2 | 10.6 | 2.1 |
| | China reports 139 new cases, | | | | | | | | |
| January 20. | including a third death. | 77.6 | 22.4 | 25.2 | 50.8 | 13.2 | 0.2 | 7.4 | 3.2 |
| | US reports first case in | | | | | | | | |
| January 21. | Washington state. | 75.4 | 24.6 | 21.7 | 38.8 | 25.2 | 0.4 | 6.9 | 6.9 |
| | Wuhan temporarily closes airport | | | | | | | | |
| | and railway stations and death toll | | | | | | | | |
| January 22. | increases to 17. | 72 | 28 | 20.8 | 28.1 | 32.4 | 0.4 | 12.7 | 5.5 |
| | WHO says virus is not a public | | | | | | | | |
| | health emergency of international | | | | | | | | |
| | concern; Beijing cancels Lunar | | | | | | | | |
| January 23 - | New Year celebrations to avoid | | | | | | | | |
| January 29 | spread | 73.2 | 26.8 | 19.5 | 35.9 | 20.1 | 0.9 | 17.4 | 6.1 |





As we show in Figure 9, the topics from this period are related to three main recognizable themes using the unsupervised clusters available on ForSight:

1) **Locations of Outbreaks**. Geographic locations such as Wuhan and China remain heavily involved in online tweets.

2) New cases. New cases are identified and places like Japan start popping up in our clusters.

3) **Death cases:** Words such as death, deadly mysterious, and infectious become more prominent in tweets discussing COVID-19.

Phase 4 (January 30 – February 2020)

From January 30 to February 20, over 15M posts on COVID-19 were shared on Twitter. The highest volume of tweets happened on January 30 and February 1 with a combined volume of 5M tweets. This high volume of tweets coincided with the news that coronavirus had been declared a global emergency by the WHO. The net sentiment shared in this period was 78.1% negative and 21.9% positive. By this time, most users had gone past the initial fear posts. The emotions shared in posts were primarily of *disgust* (32.9%), *joy* (22.4%), and *sadness* (19.7%). Emotions and sentiments are displayed below in Figure 9.

| | | Sentiment | | Emotion | | | | | |
|--------------|------------------------------|-----------|----------|---------|---------|------|----------|---------|-------|
| Date | Main Press Highlight | Negative | Positive | Fear | Disgust | Joy | Surprise | Sadness | Anger |
| January 30 - | WHO declares coronavirus a | | | | | | | | |
| February 1 | global emergency | 77.4 | 22.6 | 15.2 | 39.9 | 22.7 | 0.5 | 14.5 | 7.3 |
| February 2 - | First death reported outside | | | | | | | | |
| February 3 | China in Thailand. | 75.7 | 24.3 | 20.1 | 35.9 | 22.5 | 0.2 | 16.1 | 5.2 |

| | Japan reports 10 new cases | | 1 | | | 1 | | 1 | 1 |
|---------------|---------------------------------|------|------|------|------|------|-----|------|-----|
| February 4 - | from people aboard the cruise | | | | | | | | |
| February 6 | ship Diamond Princess. | 78.5 | 21.5 | 20.2 | 28.5 | 20.3 | 0.2 | 24.6 | 6.2 |
| | Chinese doctor who issues early | | | | | | | | |
| February 7. | warning on coronavirus dies. | 75.4 | 24.6 | 16.0 | 28.8 | 20.8 | 0.3 | 29.0 | 5.1 |
| February 8 - | First foreigner death as US | | | | | | | | |
| February 10 | citizen dies in Wuhan. | 78.8 | 21.2 | 19.7 | 28.0 | 18.5 | 0.6 | 27.2 | 6.0 |
| | WHO gives official name for | | | | | | | | |
| February 11. | coronavirus as COVID-19. | 77.8 | 22.2 | 23.4 | 27.3 | 20.5 | 3.0 | 20.1 | 5.7 |
| | Infections on the Diamond | | | | | | | | |
| February 12. | Princess rises to 175. | 79.0 | 21.0 | 31.6 | 24.1 | 22.1 | 0.7 | 16.0 | 5.5 |
| | Death toll in China reaches | | | | | | | | |
| February 13 - | 1,300 with over 60,000 | | | | | | | | |
| February 16 | infections. | 78.8 | 21.2 | 20.4 | 25.6 | 27.9 | 0.5 | 21.0 | 4.5 |
| | Quarantined Diamond Princess | | | | | | | | |
| February 17- | has the most infections outside | | | | | | | 1 | |
| February 20 | of China with 99 new cases. | 79.8 | 20.2 | 15.4 | 35.4 | 23.4 | 0.3 | 20.0 | 5.4 |

Table 3. Highlights from Phase 4

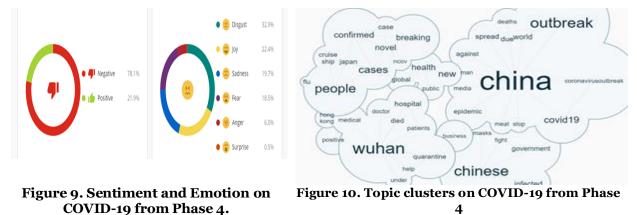


Figure 10 shows the clusters of topics formed from tweets within this period. Wuhan and Chine were still predominant in tweets discussing the disease. Other themes formed include:

1) **Global health epidemic.** The public now discussed coronavirus as a global health issues following the WHO declaration.

2) **Cruise ship:** The news by Japan that 10 people aboard the Diamond Princess, a cruise ship, have contracted the disease becomes a major topic online.

3) **New name.** Tweets now described the disease using the official name given by the WHO: COVID-19 (Corona Virus Disease 2019).

Thematic Analysis

Based on the analysis conducted, the tweets were grouped into 9 themes: *Prevention, Spread of disease, Symptoms, Food, Politics, Health organizations and media, Emotions, Humor and General Posts.* The Kappa value between the two coders in this research was 0.76, which means that the agreement among the two coders was high. For the entries with disagreements were discussed to reconcile differences in coding.

A summary of the themes and sample tweets are shown in Table 4.

Prevention: As COVID-19 is a human to human transferred diseases, tweets which discussed preventive measures such as frequently washing hands, avoiding touching eyes, nose and mouth, and wearing face mask are labeled under Prevention.

Spread of disease: There was lot of discussion on the origin of the disease. Many references were made on country of origin of COVID-19, China, and other countries which was affected by corona virus. There

were number of tweets which also discussed on Chinese population, and Chinese government and authorities.

Symptoms: There were lot of tweets which discussed COVID-19 symptoms. Identifying these symptoms at early stages and taking necessary medical attention can save lives.

Food: There were numerous tweets on food such as sea food, and pork/bacon. Some tweets mentioned halal food and vegetarian diets. There was lot of discussions and concerns on eating seafood as it was rumored to be linked to the corona virus. Some twitter users suggested others not to eat meat and was propagating wrong information, claiming eating meat causes coronavirus. The misinformation such as "eating seafood causes corona virus" calls for health authorities to disperse educational messages as soon as possible to stop spread of wrong information among users.

Politics: Tweets related to politics and political reference were captured in this theme. There were numerous tweets that referred to President Donald Trump, asking for help to stop spread of disease in U.S.A. Some twitter users who support a particular party used this outbreak to attack the opposition.

Health organizations and media: This theme includes the tweets that mentioned a media organization such as Fox news, CNN and ABC News, and also those that mentioned health organizations such as World health Organization (WHO), Center of Disease Control (CDC), and Global Health. The majority of the tweets indicates critical views towards a certain media and health organizations.

Emotions: Tweets which expressed emotions towards COVID-19 outbreak such as fear, anxiety, and anger was categorized in this theme. Corona virus outbreak has created fear among people in social media. Many tweets have shown concern, fear and anxiety related to corona virus spread. Use of vulgarity was common. Government and concerned authorities should make sure that right information is shared so as to reduce the fear among people.

Humor: Even in times of distress social media, users still found time to poke humor at events. There were several tweets containing general humor and sarcasm. There also existed some users who were simply trolling to gain clout on social media. Although most humorous tweets on such issues are generally harmless, they can sometimes be problematic as they could easily be misconstrued as racist when taken out of context. Additionally, such humor tweets could take attention away from real issues or people who need attention.

General Posts: Posts captured in this theme simply captured tweets that didn't necessarily fall in any of the categories discussed above. Most of such tweets simply shared general sentiments on the disease.

| Themes (%) | Sample Tweets |
|--|---|
| Prevention (6.8%) | "Shit just got real 🔗 Protect urself by wearing face masks and wash hands frequently" "Please make some time to read articles about this Corona Virus. We need to have serious learnings about this one bcoz its REALLY DEADLY. Indeed, Prevention is better than Cure so let's ask our gov't to ban all flights coming from these affected areas https://t.co/BpZgKOcB4q" |
| Spread of disease | "Global case tally of the novel #Wuhan #pneumonia #coronavirus = 221: ●217 in China 🌞 198 in Hubei province (Wuhan) 🏶 14 in Guangdong 🏶 5 in Beijing ●1 in Japan ●2 in Thailand ●1 South Korea" |
| (11.45%) | "Holy fuck France confirmed two cases of corona virus. And one in paris 🚱 shit is getting real now" |
| Symptoms (5.2%) | "For most patients, the #coronavirus begins and ends in their lungs, 20% end up in ICU. COVID19 patients might start out with a fever and cough that progresses to pneumonia or worse. Are hospitals around the world ready for this?#COVID #COVID19 #China #wuhan # 冠状病毒" |
| | "Symptoms of Corona Virus OFever Ocough OShortness of breath #CoronaAwareByVIJAYFans" |
| Food (7.3%) | "Chinese Muslims are safe from the corona virus just because of halal eating https://t.co/xM6oozUToM" "RT @chhapidassi Hello China!!! #NoMeat_NoCoronaVirus STOP EATING MEAT & CORONA virus come from \$\overline{4}\$ U china people can kill innocent animals and this kind of thing God is never happy. @SaintRampalJiM Now Watch satsangh (IST- 7:30pm to 8:30pm sadhna TV. https://t.co/2QDPGf2H1W" |
| Politics (2.1%) | "If we keep Republicans we get more Angry rage rallies opioid deaths corona virus pollution global warming fires iran war tariffs inflation low wages farmer bankruptcy cuts to social security klan marches in dc and lies lies lies from republicans" "@realDonaldTrump What are you doing to prepare America for the inevitable Corona virus pandemic?" |
| Health organizations and media (13.72%) | "Name changes as of yesterday - established by the World Health Organization-@WHO: #nCoV2019 virus is now SARS-CoV-2 #nCoV2019disease is now KCOVID19 #SARSCoV2 "I wasn't worried until I realized they're trying to coverup this #Wuhan outbreak. No media coverage & Twitter is actively hiding it like a MAGA tag. There's ZERO chance this isn't trending. This #Corona virus outbreak must have them terrified to be opressing free speech." |

| Emotions (9.15%) | "okay that corona virus stuff from china is pretty scary bruh!!!" "f**k corona virus all my homies hate corona virus" |
|---------------------|--|
| Humor (7.6%) | "We gotta beat the corona virus by only drinking modelo" |
| | "*someone offers me some food and I say no thank you because I dont want to take their food* "Madam, this is Africa, dont worry, there is no Corona Virus here" Okay thank you, and I proceed to take his food |
| | hahaha" |
| General Posts | "What is Corona Virus? - a thread https://t.co/Z3pA3norJn" |
| (36.68%) | "My account got suspended for 12 hours cause I told someone I hope they get the corona virus $\mathscr{O} \mathscr{O} \mathscr{O} \mathscr{O}$ " |
| | An a la sin and a second a second a second a second a factor of the second a se |

 Table 4. Sample tweets and percentage of tweets from Thematic Analysis

Conclusion and Future Implications

In this study, we have presented a comprehensive social media analysis on the spread of COVID-19 related information on Twitter. We investigated the behavioral panorama of social media users and emanant discussions around early outlook on COVID-19. Using social media analytics tool ForSight provided by Crimson Hexagon, we mined and analyzed over 40M COVID-19 related posts in a timespan starting 6 months from the first reports of the disease (December 2019 to February 2020). The study addressed two main research questions.

First, we investigated the change in sentiment and emotion overtime on COVID-19. We found that in general the net sentiment on the coronavirus was positive (72%) in the first phase of our selected timespan when the disease was non-existent. Tweets in this time period expressed joy mainly due to the happiness with research efforts towards discovering vaccines for MERS-CoV and SARS-CoV. In the second to fourth phases however, we found that sentiments were extremely negative. Phase 2 had 87% negative sentiment with the prevalent emotion being fear (61.4%). In Phase 3, there was 73.3% negative sentiment and posts displayed predominantly disgust (35.8%). In the last phase of our analysis, sentiment was 78.1% negative and 32.9% disgust emotion were uncovered in posts.

Second, we discussed the prevalent themes which inherent in posts on COVID-19. Using a thematic analysis, we grouped posts into nine themes: Prevention, Spread of disease, Symptoms, Food, Politics, Health organizations and media, Emotions, Humor and General Posts. From the health and public domain perspective, these themes will help foster a general understanding of how the public discusses disease outbreaks. For example, prevention tweets can help public health officials to gauge the public understanding of the disease when preparing various communique used for control measures. This is in addition to insights on symptoms will assist in targeted education of the public. Even further, outbreak of new symptoms can advise healthcare agencies and research labs on what to lookout for in developing appropriate measures to control the infection and spread at an early stage. We believe that social media can act as an effective educational tool for the general public and even allow twitter users to track the spread of disease. This will allow users to feel safe and take measures if a possibility exists of the disease spreading to a location close to them. These themes will help epidemiologists and corresponding agencies to arrange the required resources for preventing and fighting illness and diseases. Since our analysis did not use any advanced text mining techniques, future researchers may use such techniques to help improve the analysis and theme development in a much more granular way especially as this is an ongoing pandemic as at the time of this study.

REFERENCES

- Al Jazeera. 2020. "Timeline: How China's New Coronavirus Spread." (https://www.aljazeera.com/news/2020/01/timeline-china-coronavirus-spread-200126061554884.html, accessed February 21, 2020).
- Antonio, S. S. 2020. "Coronavirus: A Timeline of the Deadly Covid-19 Outbreak," *KABB*, February 13. (https://foxsanantonio.com/news/local/coronavirus-a-timeline-of-the-deadly-covid-19-outbreak, accessed February 21, 2020).

Boit, J., and El-Gayar, O. 2020. *Topical Mining of Malaria Using Social Media*. A *Text Mining Approach*, January 7. (https://doi.org/10.24251/HICSS.2020.466).

- Broniatowski, D. A., Dredze, M., Paul, M. J., and Dugas, A. 2015. "Using Social Media to Perform Local Influenza Surveillance in an Inner-City Hospital: A Retrospective Observational Study," *JMIR Public Health and Surveillance* (1:1). (https://doi.org/10.2196/publichealth.4472).
- CDC. 2020. "2019 Novel Coronavirus (2019-NCoV)," *Centers for Disease Control and Prevention*, , February 11. (https://www.cdc.gov/coronavirus/2019-ncov/about/symptoms.html, accessed February 16, 2020).
- CNN Editorial Research. 2020. "Wuhan Coronavirus Timeline Fast Facts," *CNN*. (https://www.cnn.com/2020/02/06/health/wuhan-coronavirus-timeline-fast-facts/index.html, accessed February 21, 2020).
- Crimson Hexagon. 2018. "Emotion Analysis: Overview," *Crimson Hexagon*, , December 10. (http://help.crimsonhexagon.com/hc/en-us/articles/211129163-Emotion-Analysis-Overview, accessed February 18, 2020).
- Crimson Hexagon. 2019a. "Forsight," *Welcome to Crimson Hexagon: A Crash Course in ForSight*, , May 8. (http://help.crimsonhexagon.com/hc/en-us/articles/115000518766-Welcome-to-Crimson-Hexagon-A-Crash-Course-in-ForSight, accessed February 18, 2020).
- Crimson Hexagon. 2019b. "Sentiment Analysis: Overview," *Crimson Hexagon*, , December 10. (http://help.crimsonhexagon.com/hc/en-us/articles/203523885-Sentiment-Analysis-Overview, accessed December 21, 2019).
- Crimson Hexagon. 2019c. "Explore Tab: Clusters," *Crimson Hexagon*, , August 18. (http://help.crimsonhexagon.com/hc/en-us/articles/202913009-Explore-Tab-Clusters, accessed December 22, 2019).
- Deiner, M. S., Lietman, T. M., McLeod, S. D., Chodosh, J., and Porco, T. C. 2016. "Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns," *JAMA Ophthalmology* (134:9), pp. 1024–1030. (https://doi.org/10.1001/jamaophthalmol.2016.2267).
- van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J. M., Wolthers, K. C., Wertheim-van Dillen, P. M. E., Kaandorp, J., Spaargaren, J., and Berkhout, B. 2004. "Identification of a New Human Coronavirus," *Nature Medicine* (10:4), pp. 368–373. (https://doi.org/10.1038/nm1024).
- Khan, A. S., Fleischauer, A., Casani, J., and Groseclose, S. L. 2010. "The Next Public Health Revolution: Public Health Information Fusion and Social Networks," *American Journal of Public Health* (100:7), pp. 1237–1242. (https://doi.org/10.2105/AJPH.2009.180489).
- Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., and Suran, M. 2015. "Detecting Themes of Public Concern: A Text Mining Analysis of the Centers for Disease Control and Prevention's Ebola Live Twitter Chat," *American Journal of Infection Control* (43:10), pp. 1109–1111. (https://doi.org/10.1016/j.ajic.2015.05.025).
- Neiger, B. L., Thackeray, R., Burton, S. H., Giraud-Carrier, C. G., and Fagen, M. C. 2012. "Evaluating Social Media's Capacity to Develop Engaged Audiences in Health Promotion Settings: Use of Twitter Metrics as a Case Study," *Health Promotion Practice*. (https://doi.org/10.1177/1524839912469378).
- Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., and Grunwald, D. 2010. "A Vision for Technology-Mediated Support for Public Participation & Assistance in Mass Emergencies & Disasters," in *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference*, ACM-BCS '10, Edinburgh, United Kingdom: BCS Learning & Development Ltd., April 14, pp. 1–12.
- Signorini, A., Segre, A. M., and Polgreen, P. M. 2011. "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic," *PLOS ONE* (6:5), p. e19467. (https://doi.org/10.1371/journal.pone.0019467).
- Sullivan, S. J., Schneiders, A. G., Cheang, C.-W., Kitto, E., Lee, H., Redhead, J., Ward, S., Ahmed, O. H., and McCrory, P. R. 2012. "What's Happening?' A Content Analysis of Concussion-Related Traffic on Twitter," *British Journal of Sports Medicine* (46:4), pp. 258–263. (https://doi.org/10.1136/bjsm.2010.080341).
- WHO.2020a."Coronavirus,"WorldHealthOrganization.(https://www.who.int/emergencies/diseases/novel-coronavirus-2019, accessed February 16, 2020).
- WHO. 2020b. "Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-NCoV),", January 30. (https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-theinternational-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novelcoronavirus-(2019-ncov), accessed February 18, 2020).