Analysis of public reactions to the novel Coronavirus (COVID-19) outbreak on Twitter

Article *in* Kybernetes · November 2020

DOI: 10.1108/K-05-2020-0258

6 authors, including:

Saleha Noor
East China University of Science and Technology

Syed Hamad Hassan Shah
Donghua University

Philippe Fournier Viger
Harbin Institute of Technology (Shenzhen)

M. Saqib Nawaz
Harbin Institute of Technology Shenzhen Graduate School

Analysis of public reactions to the novel Coronavirus (COVID-19) outbreak on Twitter

Saleha Noor and Yi Guo
School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China

Syed Hamad Hassan Shah
Glorious Sun School of Business and Management, Donghua University, Shanghai, China, and

Philippe Fournier-Viger and M. Saqib Nawaz
School of Humanities and Social Sciences, Harbin Institute of Technology (Shenzhen), Shenzhen, China

Received 2 May 2020
Revised 29 June 2020
27 July 2020
30 August 2020
Accepted 24 September 2020

Abstract

Purpose – The novel Coronavirus (COVID-19) pandemic, which started in late December 2019, has spread to more than 200 countries. As no vaccine is yet available for this pandemic, government and health agencies are taking draconian steps to contain it. This pandemic is also trending on social media, particularly on Twitter. The purpose of this study is to explore and analyze the general public reactions to the COVID-19 outbreak on Twitter.

Design/methodology/approach – This study conducts a thematic analysis of COVID-19 tweets through VOSviewer to examine people's reactions related to the COVID-19 outbreak in the world. Moreover, sequential pattern mining (SPM) techniques are used to find frequent words/patterns and their relationship in tweets.

Findings – Seven clusters (themes) were found through VOSviewer: Cluster 1 (green): public sentiments about COVID-19 in the USA. Cluster 2 (red): public sentiments about COVID-19 in Italy and Iran and a vaccine, Cluster 3 (purple): public sentiments about doomsday and science credibility. Cluster 4 (blue): public sentiments about COVID-19 in India. Cluster 5 (yellow): public sentiments about COVID-19's emergence. Cluster 6 (light blue): public sentiments about COVID-19 in the Philippines. Cluster 7 (orange): Public sentiments about COVID-19 US Intelligence Report. The most frequent words/patterns discovered with SPM were "COVID-19," "Coronavirus," "Chinese virus" and the most frequent and high confidence sequential rules were related to "Coronavirus, testing, lockdown, China and Wuhan."

Research limitations/implications – The methodology can be used to analyze the opinions/thoughts of the general public on Twitter and to categorize them accordingly. Moreover, the categories (generated by VOSviewer) can be correlated with the results obtained with pattern mining techniques.

Social implications – This study has a significant socio-economic impact as Twitter offers content posting and sharing to billions of users worldwide.

Originality/value – According to the authors' best knowledge, this may be the first study to carry out a thematic analysis of COVID-19 tweets at a glance and mining the tweets with SPM to investigate how people reacted to the COVID-19 outbreak on Twitter.

Keywords Clusters, VOSviewer, COVID-19, Sequential pattern mining, Tweets

Paper type Research paper

The authors would like to acknowledge Miss Tuba Atif for her help in data collection from Twitter. Moreover, they also acknowledge the valuable comments given by the reviewers, which have improved the quality of the paper.

Kybernetes

0368-492X
DOI 10.1108/K-05-2020-0258

> If one person thinks about something, it becomes an idea. If 1,000 people think the same thing, it becomes an evolution. If one million people think about the same thing, it becomes a revolution. If one billion people think about the same thing, it becomes a law [. . .] and this is the power of big data Shah (2020).

1. Introduction

In the past century, the world has not seen a pandemic having consequences as serious as those of the novel Coronavirus (COVID-19) pandemic. The panic, fear, uncertainty and mental stress are prevailing around the globe and the conditions are getting more and more severe. According to the World Health Organization (WHO)'s latest report (World Health Organization (WHO), 2020), more than 7.05 million people have been infected by the COVID-19, there are approximately 743,000 severe cases and 400,000 deaths. Moreover, this disease has spread to more than 200 countries, which is almost every region of the world. At the time of writing this paper, the most affected regions are the USA, Europe and Asia. Based on this assessment, the WHO has indicated that the risk level of this pandemic as very high.

Twitter has been used in the past as a prominent research tool for public health management (Breland *et al.*, 2017; Sinnenberg *et al.*, 2017). Prior studies highlighted the major role that Twitter can play in evaluating the public's opinion on pandemic virus outbreaks. For instance, Twitter was used to study the MERS-CoV virus outbreak in Saudi Arabia (Zarrad *et al.*, 2014), that of the Ebola virus (Kim *et al.*, 2015; Fung *et al.*, 2016), dengue virus in Pakistan (Kraemer *et al.*, 2018) and the global reaction to the Zika virus (Fu *et al.*, 2016; Bragazzi *et al.*, 2017). Besides, Twitter was used as an effective platform for knowledge sharing during the Zika virus outbreak affecting pregnant women from 2015 to 2016 (Lehnert *et al.*, 2017), to examine protective behaviors and risk perception of the Zika virus (Chan *et al.*, 2018) and to debunk and propagate conspiracy theories during the Zika outbreak (Wood, 2018). Moreover, the role of social media in polio prevention in India was evaluated (Kumar *et al.*, 2018). The above studies motivate us to analyze people's reactions to the COVID-19 outbreak worldwide, on Twitter.

Some of the key features of Twitter are that it is widely used to obtain real-time messages that can capture trends in epidemics and other events, as well as serve as a means for information collection and knowledge dissemination (Khan *et al.*, 2010). Twitter is known as an effective transmission medium for information and public health news which has been evident, for example, by its usage for tracking levels of disease activity and public concerns during the influenza virus subtype A H1N1 pandemic (Signorini *et al.*, 2011). Twitter messages can reveal many types of concerns, despite that the number of characters per message is restricted. Twitter can support surveillance efforts in new and innovative ways for public health (Henning, 2004). The study of Odlum and Yoon (2015) concluded that an expanded definition of public health outbreak surveillance is needed because social media content can be used to support and enhance existing early warning systems. Ueda *et al.* (2017) observed that the number of suicides increased when suicides caused a major response from Twitter users. On the other hand, there was no significant increase in suicide counts when Twitter users were not very interested in them, even though these deaths were largely covered by traditional media. In some other recent studies, Wang *et al.* (2015, 2019) demonstrated that online user text in the form of search queries can help to predict diarrhea outpatient visits. As highlighted in the above studies, analyzing online text produced by users and public opinion based on tweets can have a great impact on society.

Based on this observation, a key research question is whether interesting information could be revealed about public opinion regarding the COVID-19 pandemic by analyzing tweets collected during the first months of the outbreak. To provide an answer to that

question, we use automated data mining techniques as it allows us to analyze large amounts of data automatically. More precisely, we first perform a thematic analysis of people's reactions to COVID-19 on Twitter data (tweets) using software named VOSviewer to extract themes (clusters) based on keyword co-occurrences. Second, to discover frequent sequences of words appearing in tweets and analyze the relationships and dependencies between these words, we apply sequential pattern mining (SPM). Note that in this work, the tweets posted and shared by the general public were collected from COVID-19 related hashtags without giving importance to any geographic areas. The two main objectives of this study are:

(1) To provide a snapshot of COVID-19 related tweets during the COVID-19 outbreak to monitor trends about how information is spreading and evaluate public opinion and attitude toward the pandemic.

(2) To provide baseline data that could support future health communication related to COVID-19.

Twitter was chosen for this study for the same reasons as in previous studies (Aladwani, 2015; Nawaz *et al.*, 2017; Shahzad *et al.*, 2017):

- it draws more research because of its social impact;
- tweets are indexed by the Google search engine;
- hashtags facilitate data gathering around any event;
- tweets are easy to retrieve as an outbreak, events and news stories are generally centered around a hashtag;
- the Twitter API for data gathering is easy to use; and
- high tweets posting frequency offers up-to-the-minute analysis for an outbreak.

The rest of the paper is organized as follows. Section 2 describes the methodology used for data collection from Twitter and an overview of VOSviewer. Section 3 presents the results obtained using VOSviewer on tweets. Section 4 describes the SPM-based approach to find frequent patterns. Then, Section 5 discusses the results and research implications. Finally, Section 6 draws a conclusion and presents research directions.

2. Methodology

In this work, we followed the methodology of Yoon *et al.* (2018) to examine people's reactions on Twitter to the COVID-19 outbreak. Yoon *et al.* (2018) collected 10k tweets containing the keyword "shoes" from Twitter using Python. After cleaning the tweets, keywords were extracted from the retrieved data. Then, because there can be some inconsistencies in keyword usage, keywords were modified and corrected to obtain a unified set of terms. The result was 50,456 keywords that were used to visualize the relationships between keywords, from which only 72 unique terms were kept, which occur more than 30 times in the data set. Then, by applying the co-occurrence keyword analysis function of VOSViewer, six clusters were identified:

(1) daily shoes;

(2) shoe brands and features;

(3) shoes for kids and girls;

(4) athletic shoes;

(5) shoes as fashion accessories; and

(6) fashion combinations.

Inspired by this approach, we also applied VOSviewer on tweets related to COVID-19 and found related clusters about COVID-19. These clusters provide interesting insights into the perception of the COVID-19 outbreak by the public.

2.1 Data collection and cleaning

In this study, a tweet collection and cleaning methodology inspired by those of previous studies (Bragazzi *et al.*, 2017; Mishra *et al.*, 2017; Sinha, 2018) was used. The Twitter API and Python Tweepy library were used for tweet streaming. To collect relevant tweets, Google Trends was first searched to find frequently used Twitter hashtags about the COVID-19 pandemic. Top hashtags and keywords were coronaviruschina, coronavirus2020, coronavirusoutbreak2020, NCoV-2019, 2019-NCoV, ncov-2019, CoronaVirus2019 and COVID2019. After establishing a successful connection to Twitter, a stream of real-time tweets containing the filter words was received. Data collection was done during random periods of time every day from the January 23 to the February 7 and from the March 17 to the March 31. In total, 95,000 tweets were collected in JSON format and saved in a text file.

After tweet collection, preprocessing was applied to extract meaningful information from tweets using the Python Natural Language Tool Kit (NLTK 3.5). Tweets were cleaned by performing the following tasks:

- *Stemming*- The postfix of each word was removed (e.g. "ing" and "tion").
- *Tokenization*- The text was divided into tokens. This includes several sub-steps such as the removal of extra space, pragmatic handling like hapyyyyy as happy, gudddddd as good, etc.
- *Stop Word Removal*- Moreover, highly frequent words (stop words) were removed such as prepositions (a, an) and conjunctions (and, between).

After passing through these steps, the cleaned data were saved in a CSV file. Then, python Numpy and Pandas libraries were used to extract the top five countries and languages of these tweets. The top countries are the USA, India, Philippines, Brazil and Argentina, while the top languages are English (81,200), Spanish (8,000), Indian (2,000), Thai (1,500), French (1,000) and Portuguese (1,000). Furthermore, to remove non-English tweets, words corpus from NLTK python was used on cleaned data to drop non-English Tweets from the CSV file. This resulted in a total of 81,200 English language tweets, on which further analysis was carried out.

2.2 Application of VOSviewer for data mining

In this study, VOSviewer was used to extract clusters from English tweets about COVID-19. VOSviewer has been extensively used in bibliometric studies to discover clusters about presumption (Shah *et al.*, 2019, 2020), social media and knowledge management (Noor *et al.*, 2020a), Twitter (Noor *et al.*, 2020b), Twitter and Health Promotion (Noor *et al.*, 2020c), Openness and information technology (Vošner *et al.*, 2017) and the Internet of Things (Dabbagh *et al.*, 2020). However, VOSviewer is scarcely used as a data mining tool to explore related themes in tweets. Yoon *et al.*, 2018 used Twitter data in VOSviewer to identify users' needs and preferences in the shoes market by conducting keyword frequency and co-occurrence analyzes. They first crawled 10,000 tweets, which were pre-processed further to remove duplicate and redundant data in tweets. Keywords in the tweets were found and they conducted keyword frequency and co-occurrence analyzes using R. In last, VOSviewer was used to visualize the keyword map and they found 6 interesting clusters (themes).

VOSviewer provides distance-based visualizations of text networks and basically creates a map based on *items*. In this study, items are the keywords (terms) that occur frequently in collected tweets. A *link* provides the connection or relationship between two items (Eck and Waltman, 2009). Items and links together from a *network*. Clusters are non-overlapping, which means that an item can belong to only one cluster. Similarly, some items may not belong to any cluster. Clusters are labeled with different colors. The assignment of colors is done automatically by VOSviewer to differentiate clusters.

VOSviewer has the ability to sum-up enormous amounts of information in a single visual networking map based on the visualization of similarities (VOS) technique (Eck and Waltman, 2009). VOSviewer offers text mining functionalities for constructing co-occurrence networks of keywords extracted from English-language textual data (Waltman, van Eck and Noyons, 2010). For this, VOSviewer relies on the Apache OpenNLP toolkit (http://opennlp.apache.org) to perform part-of-speech tagging (i.e. to identify verbs, nouns, adjectives, etc). Then, VOSviewer calculates a relevance score for each noun phrase. Essentially, low scores are assigned to noun phrases if their co-occurrences with other noun phrases follow a more or less random pattern, while a high relevance score is given to noun phrases that co-occur mainly with a limited set of other noun phrases. VOSviewer can be used to visualize a co-occurrence network of these terms (Van Eck and Waltman, 2018) with high accuracy while removing manual text analysis expectation biases (Al-Barakati and Daud, 2018). This feature of VOSviewer helped us to mine the most trending tweet stories in the data set of tweets collected for this study.

3. Results and discussion

In this section, the obtained results with VOSviewer are presented and discussed.

3.1 Thematic analysis of the public reaction to Coronavirus

In this first analysis, we aimed at finding the main themes related to COVID-19 by studying keyword co-occurrences. Using VOSviewer, we first selected the option "create a map based on text data" to construct a visualization map using the tweet data. VOSviewer extracted a total of 66,268 keywords from the data set and 63 keywords out of 51,226 were selected by using a threshold value of 500 minimum number of occurrences by keyword. The visualization network of these keywords was constructed and is shown in Figure 1. These 63 keywords provide a network of themes (Figure 1), where each theme can be interpreted as a specific type of perception about COVID-19 (Zupic and Čater, 2015). A vertex's (keyword) size represents its number of occurrences in the map. The bigger a keyword is, the greater its occurrence count is. By performing this co-occurrence analysis, seven major clusters were obtained (depicted using different colors in Figure 1).

3.2 Cluster 1 (Green): public sentiments about Coronavirus in the USA

America, American person, bill, coronavirus relief bill, relief bill, Donald Trump, president, press conference, relief bill, senate democrat and *stock* were prominent keywords in this cluster, as shown in Figure 2.

This cluster shows that in the USA, public sentiments were mostly related to President Trump's government and how it handled the pandemic. The presence of the word "press conference" informs that from mid-March, the US President faced widespread scrutiny for various aspects of his administration's response to the economic and public health crisis triggered by the COVID-19. Various press conferences have been reported in tweets from President Donald Trump's side on COVID-19 briefings. In the

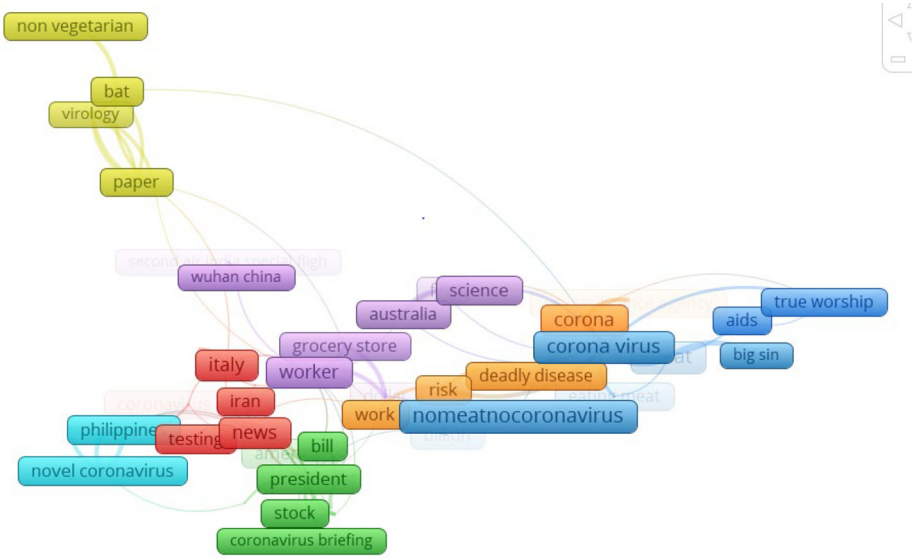

Figure 1.
Thematic analysis of COVID-19 tweets through VOSviewer

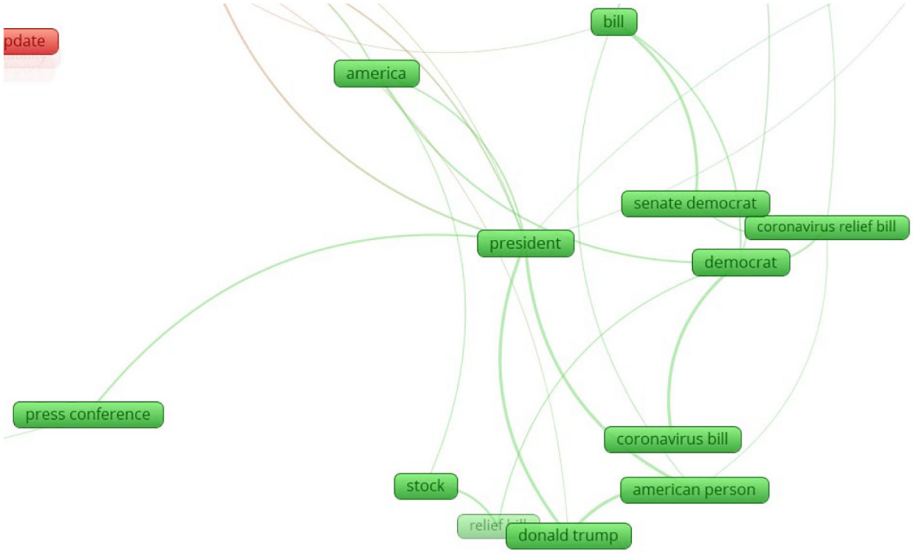

Figure 2.
Cluster 1 (green): Public sentiments about COVID-19 in the USA

cluster analysis of tweets, the most viral video was that of the March 19 press conference in which President Trump referred to COVID-19 as the "Chinese virus." Afterward, Twitter users shared a photograph purportedly taken at the event, which appeared to show that Trump was even ignoring the guidance of his aides to abstain from referring to COVID-19 as the "Chinese virus." Some sample tweets from the Donald Trump account are the following:

- The USA will be powerfully supporting those industries, like Airlines and others, that are particularly affected by the Chinese Virus. We will be stronger than ever before!
- Supporting those industries, like Airlines and others, that are particularly affected by the Chinese Virus. We will be stronger than ever before!

The presence of "bill" in this cluster is in part because of another largely retweeted story about COVID-19 in the USA. A woman in the USA said that she was billed $34,927.43 after being tested and treated for the coronavirus before President Trump signed congressional measures ensuring free diagnostic testing. Sample tweets are as follow:

- For folks who are interested in seeing what an itemized bill for COVID-19 looks like – here it is mine: $34,927.23.
- An uninsured COVID-19 patient just got her medical bill: $34,927.43.

The presence of the word "relief bill" in this cluster informed us that the relief bills signed by Trump's government have been also under discussion, after being announced from government official accounts. Some representative tweets of this cluster are:

- If you have a small business or a nonprofit, the relief bill passed by Congress offers major help for you.
- Under the financial relief bill signed on Friday, "the average worker who has lost his or her job will receive 100% of their salary for up to four full months."

3.3 Cluster 2 (red): public sentiments about Italy, Iran and a Coronavirus vaccine

In this cluster, *coronavirus death, coronavirus preventive measure, coronavirus test, coronavirus update, Iran, Italy, news, supply, testing, trump administration* and *vaccine* were prominent keywords as shown in Figure 3.

This cluster mainly put forward the theme of a sudden increase in patient number and death toll due to the COVID-19 and gave an insight into Italy and Iran's statistics about the COVID-19. The number of cases started to increase rapidly across the world and it was shown that the disease can be especially lethal for elders. Many have speculated that is why, Italy with its aged population, has been hit hard by the virus. However, some good news emerged from the grim fight against COVID-19 in that country, as a 95-year-old reportedly fully recovered. This was the most trending tweet regarding Italy. Some sample tweets are:

- Some Good News from a country close to my heart. In total, 95 years, Alma Clara Corsini is the first cured patient of Covid-19 in the province of Modena. Well done to their medical staff.
- Italy: 4032 (in 4 weeks).
- A person dies from coronavirus every 10 min in Iran

Health communities used Twitter as a platform to inform people about vaccine progresses. In our data set, a tremendous number of videos and article links were found to share information regarding vaccine development. Some sample tweets:

- #COVID19 #coronavirus has mutated into two strains. The later strain (L-type) appears far more serious than the original S-type and is now 62% of circulating cases. A challenge for vaccine developers.

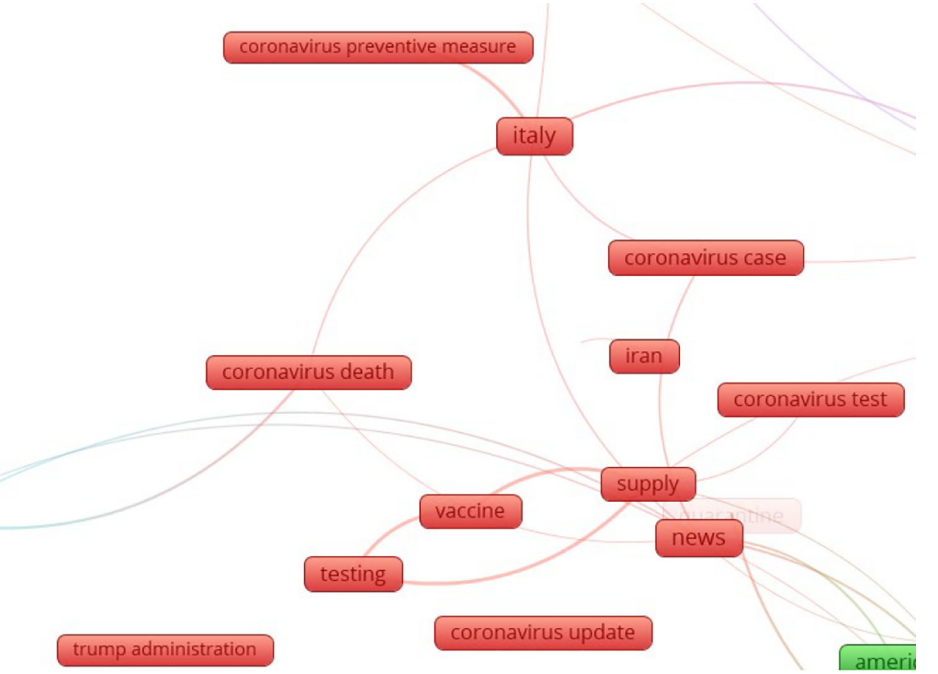

Figure 3.
Cluster 2 (red): Public sentiments about Italy, Iran and COVID-19 vaccine

- The evidence most strongly supports #2019nCoV is a vaccine strain of coronavirus either accidentally released from a laboratory or CCP performed clinical studies in humans. Human trials may make China citizens far more susceptible to ARDS upon infection.
- A #coronavirus vaccine is still many months off and even when produced will only be good for a single virus strain.
- Shanghai and Beijing Scientists say #Coronavirus has mutated into two strains:
- S-type is milder, less infectious.
- L-type (mutated later) spreads quickly, more aggressive, accounts for 70% of cases.
- WHO has announced that the first 2 candidates volunteered for COVID19 vaccines (from USA and China) are undergoing Phase 1 clinical trials, while 42 other candidates are in the pre-clinical evaluation stage.

The presence of "Trump administration" and "supply" keywords in this group let us know the drastic conditions of hospitals across the US due to the surge of critically-ill patients as the coronavirus is spreading. Short supply of masks, beds and ventilators were making it hard for medics to handle the outbreak. Many members of the medical community expressed this in the form of tweets. Some sample tweets are as follow:

- Doctors are using the same masks for a week. ICUs are jammed. Ventilators are already in short supply. The crisis is here. Where is the leadership from Washington?

- Front-line workers at New York City hospitals are struggling: "I've been in ICU care for 15 years, and this is the worst I have ever seen things."
- Police in Quebec's capital has arrested a COVID-19 patient who defied quarantine orders.
- Quarantine playlist from rolling stone.
- Blue skies were seen in China: factory shutdowns due to #coronavirus quarantine cause a steep drop in carbon emissions. Imagine-If all vehicles were #EVs we could see amazing improvements globally, lets clean up our #planet! #covid19 #climatechange #zeroemissions.

3.4 Cluster 3 (purple): public sentiments about doomsday and science credibility

In this cluster, *grocery stores, science, fire, self-isolate, Wuhan China and Australia* were prominent keywords as shown in Figure 4.

This cluster opened another interesting trending perception of Twitter's virtual communities related to doomsday, doomsday's preppers and exhibited a cynical opinion about science. The word "grocery store," appeared to be quite complicated to draw a huge concept of doomsday in tweets. With a widespread shutdown of cities and a panicky rush to buy supplies to make excessive storage of daily life goods. The shortage of toilet papers in memes and pictures of the empty aisle of supermarkets went viral on Twitter. Some sample tweets are as follows:

- Coronavirus is bringing a plague of dangerous doomsday predictions.
- As #coronavirus spreads, it fuels another dangerous contagion – doomsday prophets – that experts seem helpless to stop.
- Be very wary. Use common sense and remain calm in the midst of rampant and reckless doomsday predictions.
- When #doomsday goes down I think toilet paper will be as important as ammo. #2012.

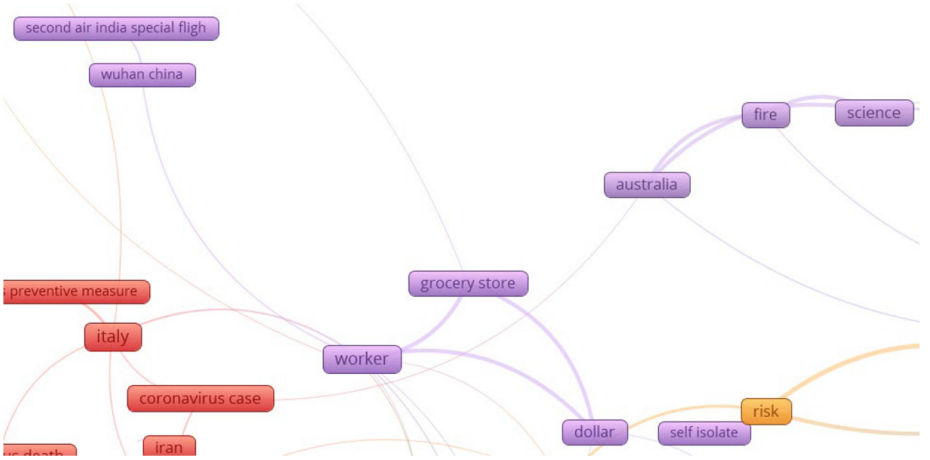

Figure 4. Cluster 3 (purple): Public sentiments about doomsday and science credibility

Some tweets also showed that some people who believed that science had failed to protect mankind and only spirituality could help. In some Twitter virtual communities, it was expressed that science had failed to stop forest fires in Australia and the spread of COVID-19 worldwide. Some sample tweets are as follows:

- Science could not stop the fire spreading in Australia, which killed millions of animals, now it is unable to stop the coronavirus spreading in China. The solution to all these disasters is spiritual knowledge.
- Science is not even worth a penny before God. Now see that there is no cure for coronavirus, no one knows how many such diseases can come in human life, to avoid this one should follow the constitution of God.

3.5 Cluster 4 (blue): public sentiments about Coronavirus in India

Big sin, coronavirus, eating meat, nomeatnocoronavirus (a hashtag) and *deadly disease* were prominent words in this cluster as shown in Figure 5.

This cluster informs us about the sentiments of the Indian public. After confirmation of the first COVID-19 case in the state of Kerala (India), a large public outrage emerged. Most of the tweets were seen exhibiting a bias related to animal meat consumption using the hashtag *#NoMeatNoCoronaVirus*. Furthermore, the spread of coronavirus presumably from animal meat (bat soup) was against some people's religious convictions or beliefs due to them being, for example, vegetarian. Some related tweets are:

- "#NoMeatNoCoronaVirus Stop eating meat China marks the deadliest day as who declare a global health emergency in the fight against Wuhan coronavirus."
- "#NoMeatNoCoronaVirus Stop eating meat."
- "#NoMeatNoCoronaVirus Eating meat causes fatal diseases, so diseases such as cancer, AIDS, troubles us."

3.6 Cluster 5 (yellow): public sentiments about Coronavirus emergence

In this cluster, *Wuhan, bat, virologist* and *paper* were prominent keywords as shown in Figure 6.

This cluster provides information about what some people thought were reasons for the COVID-19 outbreak. For example, an article named "Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confers variable susceptibility to SARS-CoV entry" was retweeted several times (3,200 times with discussion) to promote the theory that COVID-19 could have been engineered by researchers of the Wuhan Institute of Virology. A survey

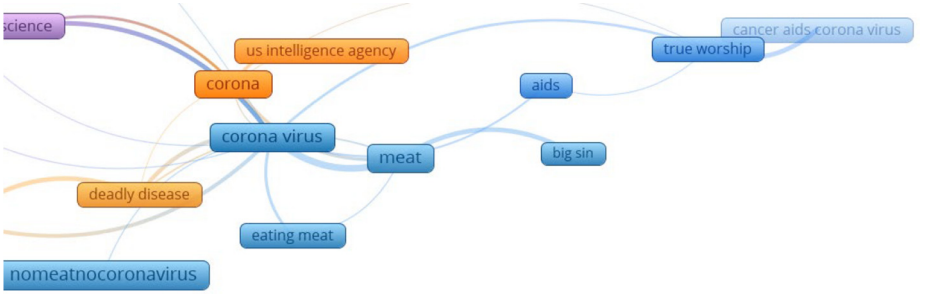

Figure 5.
Cluster 4 (blue): Public sentiments about COVID-19 in India

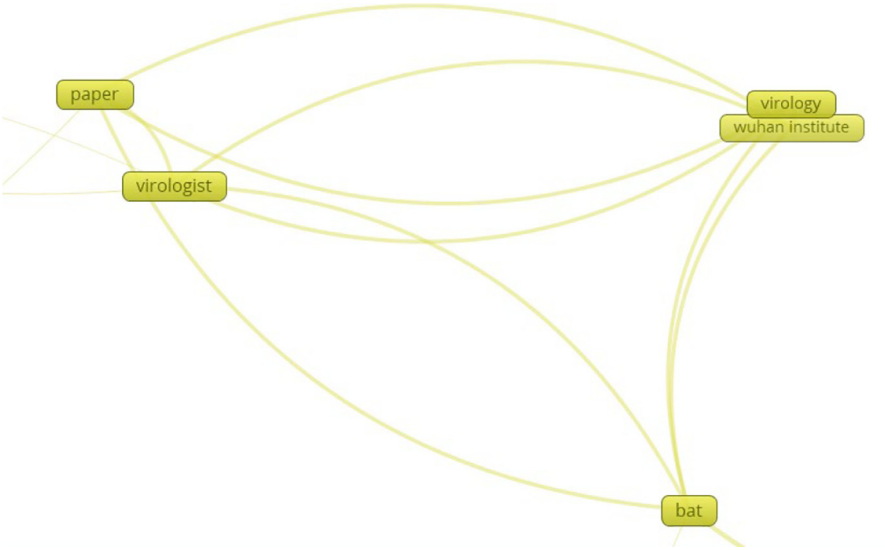

Figure 6.
Cluster 5 (yellow): Public sentiments about COVID-19 emergence

conducted on Twitter found that 11.80% of people believed that COVID-19 spread from a seafood market, while 88.20% believed that it was engineered. Some related tweets are:

- "#Coronavirus may have originated in lab linked to China's self-prepared program."
- "#Breaking: 1. Phased progress has been made in the #traceability of the #coronavirus from #China. From 585 environmental samples from the Huanan #SeafoodMarket in #Wuhan for the first time, 33 samples were detected to contain #novelcoronavirus #nucleicacids."

3.7 Cluster 6 (light blue): public sentiments about Coronavirus in the Philippines

In this cluster, *first death, novel coronavirus, Philippines* and *second case* were prominent keywords as shown in Figure 7.

This cluster informed us about people's emotions regarding the first death due to COVID-19 in the Philippines. The 44-year-old man was the second confirmed case, a companion of the woman who was the first confirmed case. Both people were from Wuhan, China – the city at the epicenter of the coronavirus outbreak. A wave of fear was observed through tweets from the Philippines and neighboring countries. This outbreak shook not only the whole world but also had an influence on how some countries managed their borders. For example, the Philippines temporarily banned non-Filipino travelers arriving from mainland China, Hong Kong and Macau. Overall, this cluster has shown a sentiment from different corners of the world toward boycotting travels to or from China. Some tweets are:

- "The country's first death of a patient who tested positive for novel coronavirus, a 44-year-old man who was the partner of the first confirmed case in PH."
- "Hong Kong hospital workers are threatening to strike against coronavirus outbreak if the government doesn't close borders with China."
- "OMG Everyone please stop transiting in Wuhan and even in China! Please take it seriously."

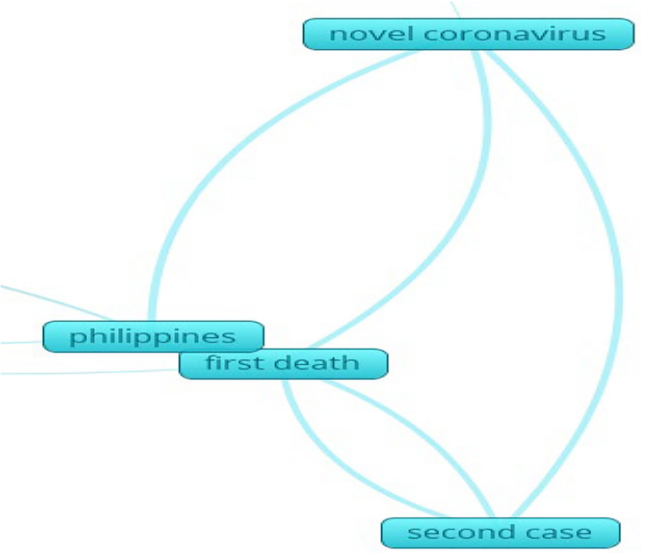

Figure 7.
Cluster 6 (light blue): Public sentiments about COVID-19 in the Philippines

3.8 Cluster 7 (orange): public sentiments about Coronavirus US intelligence report

In this cluster *deadly disease, risk, work* and *USA intelligence agency* were prominent keywords as shown in Figure 8.

As the number of COVID-19 patients was increasing dramatically in the USA (March 17: 10,000 cases, March 23: more than 50,000, March 26: more than 100,000, March 29: 150,000, March 31: more than 200,000). This cluster informs us that American agencies warned the government about the outbreak but no strict precautionary measures were taken. Some sample tweets are as follow:

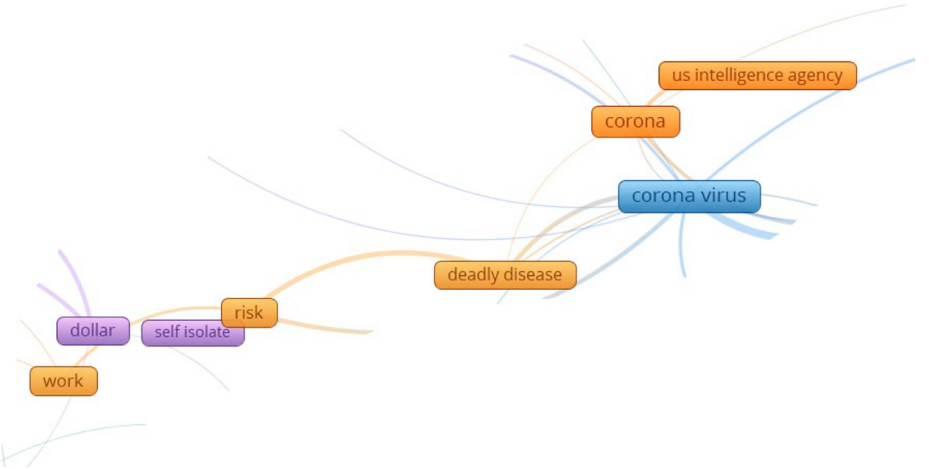

Figure 8.
Cluster 7 (orange): Public sentiments about COVID-19 US Intelligence Report

- Quite possibly a lot, and for quite a while: As reported in a @washingtonpost last week (see below) Intelligence agencies "were issuing ominous, classified warnings in January and February while President Trump and lawmakers played down the threat."
- Yes even as late as February the Trump administration wasn't taking it seriously. I think Trump and his supporters are banking on mass amnesia that forgets how long this problem was ignored.
- The intelligence reports didn't predict when the virus might land on US shores or recommend particular steps that public health officials should take, issues outside the purview of the intelligence agencies. However, they did track the spread of the virus in China and later in other countries and warned that Chinese officials appeared to be minimizing the severity of the outbreak.

4. Mining frequent patterns in tweets about Coronavirus

In the field of data mining, SPM techniques (Fournier-Viger *et al.*, 2017) are often used to find interesting and useful patterns (information) that are hidden in large corpora of sequential data. In this section, SPM techniques are used to analyze collected tweets for discovering common words/patterns and hidden relationships between them.

For this analysis, the corpus of collected tweets was converted into a suitable format to apply SPM techniques on them. Each line in the corpus, called tweets corpus (*TC*), represents a tweet, which is basically a sequence of words. In SPM, various measures are used to investigate the importance and interestingness of sequences in a sequence database. The most famous one is the Support measure. The Support of a sequence S_α in the *TC* is the total number of sequences (S) that contain S_α and is formally defined as:

$$\sup(S_\alpha) = |\ \{S\ |S_\alpha\ \ \sqsubseteq S\ \ \wedge\ \ S\ \ \in TC\}|$$

A sequence, S, is a *frequent sequence* (also called a *sequential pattern*), iff *sup(S)* $\geq$ *minsup*, where *minsup* (minimum support) is a threshold set by the user.

The SPMF data mining library (Fournier-Viger *et al.*, 2016) was used to analyze the *TC*. SPMF is an open-source JAVA library that is specialized in pattern mining tasks. It offers implementations of more than 196 data mining algorithms. It is fast, provides both command-line and graphical interfaces, and lightweight with no dependencies on other libraries.

The SPMF implementation of the Apriori algorithm (Agrawal and Srikant, 1994) was first applied to the corpus to reveal frequent sets of words. A particularity of Apriori is that it does not consider the order of words in tweets. In the context of this study, Apriori takes a corpus and the *minsup* threshold as input and returns frequent sets of words as output. Table 1 lists the most frequent patterns (sets of words) discovered in the *TC* by the Apriori algorithm. For a *minsup* threshold of 10%, Apriori generated 392 frequent itemsets (frequent sets of words). Note that there are some frequent terms such as WHO which were not included in this analysis because the English word "who" also occurred in the corpus alongside the acronym WHO. Most of the terms found in VOSviewer clusters were found in the frequent words/patterns identified by the Apriori algorithm.

Then, we tested some SPM mining algorithms that consider the order of words (unlike Apriori) on the corpus. Several fast and memory-efficient SPM algorithms have been

proposed in recent years. The TKS (Top-k Sequential) algorithm finds the top-k sequential patterns in sequences (Fournier-Viger *et al.*, 2013), where k is a user-defined parameter that represents the number of sequential patterns to be discovered by the algorithm. The parameter k is used in place of the *minsup* threshold because setting it is more intuitive than setting and fine-tuning the *minsup* support. The parameter k puts an upper bound on the total number of patterns to be discovered by the algorithm. Unlike TKS, the CM-SPAM algorithm offers the *minsup* threshold (Fournier-Viger *et al.*, 2014). Results obtained by TKS and CM-SPAM are listed in Table 2. It is important to point out that Apriori finds patterns with high support values, whereas the support values for the same patterns are comparatively low for TKS and CM-SPAM. This is because this latter consider a strict ordering between words. Patterns discovered by TKS and CM-SPAM are similar. Moreover, the total number of patterns explored by TKS depends on the value of k. CM-SPAM explored many more patterns than TKS and fewer than Apriori. CM-SPAM generated more frequent patterns than Apriori.

But sequential patterns that appear frequently in a corpus with low confidence are not useful for decision-making or prediction. Thus, we considered extracting sequential rules, a type of sequential pattern that is found by considering not only their support but also their confidence. A sequential rule of form $X \Rightarrow Y$ indicates a sequential relationship between two unordered sets of words (X and Y) that appear in the same tweet. The rule r: $X \Rightarrow Y$ means that if words in X occur in a tweet, the words of Y will occur afterward in the same tweet. The confidence and support of a rule r in TC are defined as:

$$conf_{TC}(\mathrm{r}) = \frac{|\{S|r \sqsubseteq S \ \wedge S \ \in TC\}|}{|\{S|X \sqsubseteq S \wedge S \ \in \ TC\}|}$$

$$sup_{TC}(\mathrm{r}) = \frac{|\{S|r \sqsubseteq S \ \wedge \ S \ \in \ TC\}|}{|TC|}$$

A rule r is called a *frequent sequential rule* iff sup_{TC} (r) $\geq$ *minsup* and r is called a *valid sequential rule* if and only if it is frequent and $conf_{TC}$ (r) $\geq$ *minconf*, where the thresholds (*minsup, minconf* $\in [0, 1]$) are set by the user.

The task of mining sequential rules in a corpus consists of finding all the valid sequential rules. For this, the equivalence class-based sequential rule miner (ERMiner) algorithm was used (Fournier-Viger *et al.*, 2014), as it is one of the most efficient algorithms for this task. To apply ERMiner, the confidence (*misconf*) threshold was set to 5%, which means that rules must have the confidence of at least *5*% (a rule $X => Y$ has a confidence of 5% if the set of words in X is followed by the set of words in Y at least 5% of the times when X appears in a tweet). For example, the first rule in Table 3 shows that *12*% of the time, the word is lockdown followed by a *coronavirus*.

Table 1. Discovered frequent words with apriori

Frequent words/patterns	Support	Frequent sets of words	Support
Covid19	6,396	Trump	1,166
Coronavirus	9,612	India	218
Testing	549	Italy	411
Pandemic	773	lockdown	281
Chinese Virus	175	Wuhan Virus	86

To the best of our knowledge, no published study has used SPM techniques for the analysis of COVID-19 related tweets. Two recent studies (Mohamadou *et al.*, 2020; Shi *et al.*, 2020) provided a comprehensive review of how mathematical models and artificial intelligence techniques have been used for the forecasting, diagnosis, detection and prediction of the COVID-19 disease. However, most of the studies considered COVID-19 medical imaging data (such as computed tomography (CT) and X-ray). For example, Apostolopoulos and Mpesiana (2020), Butt *et al.* (2020) and Ozturk *et al.* (2020) applied neural networks on CT scans and X-Ray by using supervised learning techniques such as Support vector machines (SVM), whereas Barstugan *et al.* (2020) and Li *et al.* (2020) applied logistic regression (LR).

5. Discussion

Overall, a huge socio-economic change has been observed throughout the world due to COVID-19. The people are affected quite differently depending on their locations and their business. This study thematic analysis shows that common people have diverse views about COVID-19 based on religious, socio-economic and cultural backgrounds. Despite studying only English language tweets, the predominance of Spanish, Portuguese and French tweets reflected panic and pain about the sudden exponential growth of COVID-19 in European countries. The rise of COVID-19 related Tweets in English, in mid and late March, indicated that people in the USA were concerned about the shortage of daily use goods, food and health safety. Doomsday's preppers were active on Twitter, sending messages about preparing well for the world's end. The Government of the USA was active on social media to calm down common folks by announcing the world's largest relief package. Furthermore, the WHO's used Twitter to publish health guidelines. Therefore, local health agencies and public health practitioners may also investigate the social and behavioral barriers to COVID-19 infection control and information.

TKS	Sup	CM-SPAM	Sup
Covid19	5,120	Covid19	5,120
Coronavirus	6,750	Coronavirus	6,750
Testing	439	Testing	439
Pandemic	533	Pandemic	533
Chinese virus	95	Chinese virus	95
Trump	806	Trump	806
India	146	India	146
Italy	244	Italy	244
lockdown	193	lockdown	193
Wuhan virus	25	Wuhan virus	25

Table 2. Discovered frequent words/patterns with TKS and CM-SPAM

Words/patterns	SUP	Confidence (%)
Coronavirus ⇒ lockdown	15	12
Chinese ⇒ Virus	25	8
Coronavirus ⇒ testing	45	8
Originated in ⇒ China, Wuhan	23	6
Coronavirus ⇒ COVID-19	87	5

Table 3. Discovered sequential rules discovered between words/patterns

The interesting question may be raised in the reader's mind, whether there is any cultural bias in this study. In terms of data collection, the tweets analyzed in this paper were not collected based on any geographic boundaries. Thus, the collected tweets represent various areas around the world. The reason is that the main goal of this study was to analyze the most retweeted stories related to COVID-19 to evaluate the general public reaction on Twitter that is not tied to a specific country. It should be noted that although cluster names mentioned geographic areas, data collection was not done to target these areas. The cluster names were given after finding the clusters by observing that retweeted stories were more popular in some specific countries. In future work, it could be an interesting avenue to collect tweets from specific countries or areas purposely and focus on specific languages to perform a more detailed analysis of the local reaction of people in these regions.

The origin of COVID-19 was discussed on the Twitter platform, some declaring it as a bioweapon, while some blaming it on meat. The presence of embedded URL links suggested that user-generated content remains the preferred direct channel to send COVID-19 information by Twitter users worldwide. Throughout the world, folks were found to blame their governments for being unable to control the spread of COVID-19. Therefore, government agencies should focus on providing a trustworthy environment to lessen the stress on people and reduce uncertainty related to COVID-19. Twitter resonated well with the public concerns about COVID-19 vaccination. The Lockdown brought a different living style to people's life, some sharing sad music lists, women sharing videos and pictures of household chores and preparing food and most of the tweets expressed people badly missing their work routines and going out. Furthermore, there was a lot of negativity associated with COVID-19, but the most retweeted positive story related to COVID-19 was about climate change. Countries that have been under stringent lockdowns to stop the spread of the coronavirus have experienced an unintended benefit. The outbreak has, at least in part, contributed to a noticeable drop in pollution and greenhouse gas emissions in some countries.

The main theoretical contribution of this study is the application of VOSviewer for the thematic analysis of tweets related to the new coronavirus (COVID-19) and the application of SPM to discover in its frequent patterns. In prior studies, VOSviewer and SPM algorithms have been applied separately but in this study, we have applied both techniques simultaneously on COVID-19 tweets data. Moreover, VOSviewer has been extensively used in bibliometric analyzes, but not to explore related themes from original tweets as a data mining tool except by Yoon *et al.* (2018), according to our best knowledge. Yoon *et al.* (2018) analyzed electronic word-of-mouth from 10,000 tweets related to shoe characteristics and found six clusters based on keyword co-occurrence analysis. Similarly, we applied VOSviewer on English tweets related to the coronavirus (COVID-19) and found seven clusters related to the public reaction to COVID-19. By applying SPM algorithms, we found some interesting word patterns. However, finding frequent patterns and sequential rules with SPM has some limitations.

First, SPM is an enumeration problem where the aim is to find all the frequent subsequences in the corpus. A sequence containing n items (keywords in this work) in a corpus can have up to $2^n - 1$ distinct subsequences. For example, if a tweet contains 8 keywords, then there are $2^8 - 1 = 255$ possible sub-tweets. This means that the size of the search space (that is the total number of possible sub-tweets) is very large even if there are few tweets in the corpus. In fact, the overall size of the search space also depends on how similar the tweets are in the corpus and on whether the k and *minsup* values are low or high. Second, the presence of redundant keywords (such as if, then, is,

the and of) have a direct effect on discovered frequent patterns and sequential rules. For example, unimportant frequent words such as "the," "of" "in the," "there is," "they do not care" have high support (for some in thousands). The sequential rules between such frequent words have higher confidence than those of important frequent sets of words. This was also evident with preliminary obtained results obtained in this study, which indicated that the total number of keywords in each tweet has a direct correlation with the efficiency of SPM algorithms.

The emergence and origin of COVID-19 have received diverse opinions on Twitter. Obtained results and discussion related to this controversial topic in this study is not the author's own point of view. Rather it is the opinion of the general public that was evaluated from tweets with VOSviewer and SPM techniques. It is also important to mention here that this study used very time-specific tweets that were collected within two months of the COVID-19 outbreak. This data is corrected enough to support an initial perception of the general public about the COVID-19. We anticipate that results might have been different if more tweets were included in the data set. Moreover, due to the time sensitiveness of tweets, people's perception might change with the passage of time. Thus, their reaction and opinions should be interpreted with caution.

6. Conclusion, limitations and future direction

This study has shown that Twitter can be used to gain interesting insights into the general public opinion about the COVID-19 outbreak. It was observed that several users were sending tweets to seek information related to COVID-19. Moreover, several interesting perspectives were detected in tweets such as opinions about the high cost of COVID-19 treatments, societal changes, the recovery of a 90 years old woman and doomsday. Tweets by the WHO have also shown a quick and organized public health response to create awareness about COVID-19. The clustering done in this study provided data that could be used to support and enhance existing early warning systems and for further research in academia related to COVID-19.

This study is time-sensitive i.e. tweets were collected in February-March 2020 but are valid enough to perform thematic analysis of the public reaction to COVID-19. Seven clusters (themes) emerged from the keyword co-occurrence analysis of tweets using VOSviewer, which highlighted the current conditions of COVID-19. Moreover, frequent patterns of words were discovered from tweets using SPM techniques. We found that keywords in tweets through which VOSviewer generated clusters were in fact frequent patterns obtained by SPM, which means that results of VOSviewer and SPM correlated with each other.

There are several directions for future work. First, other researchers may analyze the Twitter data with VOSviewer and SPM on a larger data set that contains millions of tweets. We anticipate that results may be slightly different on a bigger data set, which may reveal additional insights. Second, only Twitter data was used in this study. It would be interesting to extract and analyze data from other online platforms (such as Google Trend, Google News, Facebook and YouTube) to evaluate the correlation with information found in tweets to see if the results of this study are consistent with those for other online platforms. Third, other pattern mining techniques could be applied such as high utility itemset mining (Fournier-Viger *et al.*, 2019), where the focus is to find patterns that have high importance based on individual weights assigned to each word.

References

Agrawal, R. and Srikant, R. (1994), "Fast algorithms for mining association rules in large databases", *Proc. of the 20th International Conference on Very Large Data Bases (VLDB'94)*, pp. 487-499.

Aladwani, A.M. (2015), "Facilitators, characteristics, and impacts of Twitter use: Theoretical analysis and empirical illustration", *International Journal of Information Management*, Vol. 35 No. 1, pp. 15-25, doi: 10.1016/j.ijinfomgt.2014.09.003.

Al-Barakati, A. and Daud, A. (2018), "Venue-Influence language models for expert finding in bibliometric networks", *International Journal on Semantic Web and Information Systems*, Vol. 14 No. 3, pp. 184-201, doi: 10.4018/IJSWIS.2018070109.

Apostolopoulos, I.D. and Mpesiana, T.A. (2020), "Covid-19: automatic detection from Xray images utilizing transfer learning with convolutional neural networks", *Physical and Engineering Sciences in Medicine*, Vol. 43 No. 2, pp. 635-640, doi: 10.1007/s13246-020-00865-4.

Barstugan, M. Ozkaya, U. and Ozturk, S. (2020), "Coronavirus (COVID-19) classification using CT images by machine learning methods", CoRR, abs/2003.09424.

Bragazzi, N.L., Alicino, C., Trucchi, C., Paganino, C., Barberis, I., Martini, M., Sticchi, L., Trinka, E., Brigo, F., Ansaldi, F. and Icardi, G. (2017), "Global reaction to the recent outbreaks of Zika virus: insights from a big data analysis", in Donald, R. (Ed.), *Plos One*, Vol. 12 No. 9, p. e0185263, doi: 10.1371/journal.pone.0185263.

Breland, J.Y., Quintiliani, L.M., Schneider, K.L., May, C.N. and Pagoto, S. (2017), "Social media as a tool to increase the impact of public health research", *American Journal of Public Health*, Vol. 107 No. 12, doi: 10.2105/AJPH.2017.304098.

Butt, C., Gill, J., Chun, D. and Babu, B.A. (2020), "Deep learning system to screen coronavirus disease 2019 pneumonia", *Applied Intelligence*, doi: 10.1007/s10489-020-01714-3.

Chan, M.p.S., Winneg, K., Hawkins, L., Farhadloo, M., Jamieson, K.H. and Albarracín, D. (2018), "Legacy and social media respectively influence risk perceptions and protective behaviors during emerging health threats: a multi-wave analysis of communications on zika virus cases", *Social Science and Medicine*, doi: 10.1016/j.socscimed.2018.07.007.

Dabbagh, M., Kakavand, M. and Tahir, M. (2020), "Towards integration of blockchain and IoT: a bibliometric analysis of state-of-the-Art", *Advances in Intelligent Systems and Computing*, pp. 27-35, doi: 10.1007/978-3-030-23813-1_4.

Eck, N.J.V. and Waltman, L. (2009), "How to normalize cooccurrence data? An analysis of some well-known similarity measures", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 8, pp. 1635-1651, doi: 10.1002/asi.21075.

Fournier-Viger, P., *et al.* (2019), "'A survey of high utility itemset mining", in Fournier-Viger, P., Lin, J. W., Nkambou, R., Vo, B., Tseng, V. (Eds) *High-Utility Pattern Mining. Studies in Big Data*, Vol. 51. Springer, 10.1007/978-3-030-04921-8_1.

Fournier-Viger, P., Gueniche, T., Zida, S. and Tseng, V.S. (2014), "ERMiner: Sequential rule mining using equivalence classes", *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, doi: 10.1007/978-3-319-12571-8.

Fournier-Viger, P., Gomariz, A., Campos, M. and Rincy, T. (2014), "Fast vertical mining of sequential patterns using co-occurrence information", *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 40-52, doi: 10.1007/978-3-319-06608-0_4.

Fournier-Viger, P., Jerry, C.W.L., Rage, U.K., Yun, S.K. and Rincy, T. (2017), "A survey of sequential pattern mining. Data science and pattern recognition", *Ubiquitous International*, Vol. 1 No. 1, pp. 54-77.

Fournier-Viger, P., Gomariz, A., Gueniche, T., Mwamikazi, E. and Thomas, R. (2013), "TKS: efficient mining of top-k sequential patterns", in *Lecture Notes in Computer Science (Including Subseries*

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), doi: 10.1007/978-3-642-53914-5_10.

Fournier-Viger, P., Lin, J.C.W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z. and Lam, H.T. (2016), "The SPMF open-source data mining library version ", *In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, p. 2, doi: 10.1007/978-3-319-46131-1_8.

Fu, K.W., Liang, H., Saroha, N., Tse, Z.T.H., Ip, P. and Fung, I.C.H. (2016), "How people react to zika virus outbreaks on twitter? A computational content analysis", *American Journal of Infection Control*, Vol. 44 No. 12, doi: 10.1016/j.ajic.2016.04.253.

Fung, I.C.H., Duke, C.H., Finch, K.C., Snook, K.R., Tseng, P.L., Hernandez, A.C., Gambhir, M., Fu, K.W. and Tse, Z.T.H. (2016), "Ebola virus disease and social media: a systematic review", *American Journal of Infection Control*, Vol. 44 No. 12, doi: 10.1016/j.ajic.2016.05.011.

Henning, K.J. (2004), "What is syndromic surveillance?", *MMWR. Morbidity and Mortality Weekly Report*, Vol. 53.

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*, Vol. 100 No. 7, doi: 10.2105/AJPH.2009.180489.

Kim, E.H.J., Jeong, Y.K., Kim, Y., Kang, K.Y. and Song, M. (2015), "Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news", *Journal of Information Science*, Vol. 42 No. 6, doi: 10.1177/0165551515608733.

Kraemer, M.U.G., Bisanzio, D., Reiner, R.C., Zakar, R., Hawkins, J.B., Freifeld, C.C., Smith, D.L., Hay, S.I., Brownstein, J.S. and Alex Perkins, T. (2018), "Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan", *EPJ Data Science*, Vol. 7 No. 1, doi: 10.1140/epjds/s13688-018-0144-x.

Kumar, A., *et al.* (2018), "Investigating the role of social media in polio prevention in India: a Delphi-DEMATEL approach", *Kybernetes*, Vol. 47 No. 5, pp. 1053-1072, doi: 10.1108/K-06-2017-0210.

Lehnert, J.D., Ellingson, M.K., Goryoka, G.W., Kasturi, R., Maier, E. and Chamberlain, A.T. (2017), "Use of obstetric practice web sites to distribute Zika Virus information to pregnant women during a Zika Virus outbreak", *Journal of Public Health Management and Practice*, doi: 10.1097/PHH.0000000000000537.

Li, K., Wu, J., Wu, F., Guo, D., Chen, L., Fang, Z. and Li, C. (2020), "The clinical and chest CT features associated with severe and critical COVID-19 pneumonia", *Investigative Radiology*, Vol. 55 No. 6, pp. 327-331, doi: 10.1097/RLI.0000000000000672.

Mishra, P., Rajnish, R. and Kumar, P.N. (2017), "Sentiment analysis of twitter data: Case study on digital India", *2016 International Conference on Information Technology, InCITe 2016 – The Next Generation IT Summit on the Theme – Internet of Things: Connect your Worlds*, 10.1109/INCITE.2016.7857607.

Mohamadou, Y., Halidou, A. and Kapen, P.T. (2020), "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID", *Applied Intelligence*, Vol. 50 No. 11, p. 19, doi: 10.1007/s10489-020-01770-9.

Nawaz, M.S., Bilal, M., Lali, M.I., Ul Mustafa, R., Aslam, W. and Jajja, S. (2017), "Effectiveness of social media data in healthcare communication", *Journal of Medical Imaging and Health Informatics*, Vol. 7 No. 6, pp. 1365-1371, doi: 10.1166/jmihi.2017.2148.

Noor, S., Guo, Y., Shah, S.H.H., *et al.* (2020a), "Bibliometric analysis of social media as a platform for knowledge management", *International Journal of Knowledge Management*, Vol. 16 No. 3, pp. 33-51, doi: 10.4018/IJKM.2020070103.

Noor, S., Guo, Y., Shah, S.H.H., *et al.* (2020b), "Research synthesis and thematic analysis of twitter through bibliometric analysis", *International Journal of Semantic Web and Information Systems*, Vol. 16 No. 3, pp. 88-109, doi: 10.4018/IJSWIS.2020070106.

Noor, S., Guo, Y., Shah, S.H.H. and Halepoto, H. (2020c), "Bibliometric analysis of Twitter knowledge management publications related to health promotion", *International Journal of Semantic Web and Information Systems*, in Li, G., Shen, H.T., Yuan, Y., Wang, X., Liu, H. and Zhao, X. (Eds), Springer International Publishing, Cham, pp. 341-354.

Odlum, M. and Yoon, S. (2015), "What can we learn about the Ebola outbreak from tweets?", *American Journal of Infection Control*, Vol. 43 No. 6, pp. 563-571, doi: 10.1016/j.ajic.2015.02.023.

Ozturk, T., *et al.* (2020), "Automated detection of covid-19 cases using deep neural networks with X-ray images", *Computers in Biology and Medicine*, Vol. 121, pp. 103792, doi: 10.1016/j.compbiomed.2020.103792.

Shah, S.H.H., *et al.* (2019), "Prosumption: bibliometric analysis using HistCite and VOSviewer", *Kybernetes*, pp. 1-24, doi: 10.1108/K-12-2018-0696.

Shah, S.H.H., *et al.* (2020), "Research synthesis and new directions of Prosumption: a bibliometric analysis", *International Journal of Information and Management Sciences*, Vol. 31 No. 1, pp. 79-98, doi: 10.6186/IJIMS.20200331(1).0005.

Shahzad, B., *et al.* (2017), "Discovery and classification of user interests on social media", *Information Discovery and Delivery*, Vol. 45 No. 3, pp. 130-138, doi: 10.1108/IDD-03-2017-0023.

Shi, F., *et al.* (2020), "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19", *IEEE Reviews in Biomedical Engineering*, doi: 10.1109/RBME.2020.2987975.

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 US during the influenza a H1N1 pandemic", *PLoS One*, Vol. 6 No. 5, doi: 10.1371/journal.pone.0019467.

Sinha, A.K. (2018), "Sentiment analysis of twitter data: case study on digital India Prerna", *in*, pp. 170-184, doi: 10.4018/978-1-5225-3787-8.ch011.

Sinnenberg, L., *et al.* (2017), "Twitter as a tool for health research: a systematic review", *American Journal of Public Health*, Vol. 107 No. 1, pp. E1-E8, doi: 10.2105/AJPH.2016.303512.

Ueda, M., *et al.* (2017), "Tweeting celebrity suicides: users' reaction to prominent suicide deaths on twitter and subsequent increases in actual suicides", *Social Science and Medicine*, Vol. 189, pp. 158-166, doi: 10.1016/j.socscimed.2017.06.032.

Van Eck, N.J. and Waltman, L. (2018), "VOSviewer manual", *Universitteit Leiden*, doi: 10.3402/jac.v8.30072.

Vošner, H.B., *et al.* (2017), "Openness and information technology: a bibliometric analysis of literature production", *Kybernetes*, Vol. 46 No. 5, pp. 750-766, doi: 10.1108/K-10-2016-0292.

Waltman, L., van Eck, N.J. and Noyons, E.C.M. (2010), "A unified approach to mapping and clustering of bibliometric networks", *Journal of Informetrics*, Vol. 4 No. 4, pp. 629-635, doi: 10.1016/j.joi.2010.07.002.

Wang, Y., *et al.* (2015), "Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion", *Knowledge-Based Systems*, Vol. 88, doi: 10.1016/j.knosys.2015.08.013.

Wang, Z., *et al.* (2019), "TDDF: HFMD outpatients prediction based on time series decomposition and heterogenous data fusion in Xiamen, China", *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, doi: 10.1007/978-3-030-35231-8_48.

Wood, M.J. (2018), "Propagating and debunking conspiracy theories on Twitter during the 2015-2016 Zika Virus outbreak", *Cyberpsychology, Behavior, and Social Networking*, Vol. 21 No. 8, doi: 10.1089/cyber.2017.0669.

World Health Organization (WHO) (2020), "Novel Coronavirus (2019-nCoV) Situation report- 103", WHO Bulletin, Geneva.

Yoon, Y.S., Oh, H.W. and Park, K. (2018), "Mining twitter to identify customers' requirements and shoe market segmentation", *9th International Conference on Information and Communication*

Technology Convergence: ICT Convergence Powered by Smart Intelligence, ICTC 2018, 10.1109/ICTC.2018.8539363.

Zarrad, A., Jaloud, A. and Alsmadi, I. (2014), "The evaluation of the public opinion - a case study: MERS-CoV infection virus in KSA", *Proceedings - 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, UCC 2014*, 10.1109/UCC.2014.107.

Zupic, I. and Čater, T. (2015), "Bibliometric methods in management and organization", *Organizational Research Methods*, Vol. 18 No. 3, pp. 429-472, doi: 10.1177/1094428114562629.

Corresponding author
Yi Guo can be contacted at: guoyi@ecust.edu.cn