Trend of social media news : a viewpoint of COVID-19 tweets using natural language processing

Olaleye, Sunday Adewale

International Business Information Management Association (IBIMA)

https://erepo.uef.fi/handle/123456789/24198

Downloaded from University of Eastern Finland's eRepository
Trend of Social Media News: A Viewpoint of COVID-19 Tweets Using Natural Language Processing

Sunday Adewale OLALEYE*
Department of Marketing, Management and International Business, Oulu Business School, Erkki Koiso-Kantilan Katu, Oulu 90570, Finland, sunday.olaleye@oulu.fi*

Oluwafemi Samson BALOGUN
School of Computing, University of Eastern Finland, FI-70211, Finland, samson.balogun@uef.fi

Bukola SALAMI
School of Computing, University of Eastern Finland, FI-70211, Finland, bukolas@uef.fi

Abstract
The meteoric rise of social media news during the ongoing COVID-19 is worthy of advanced research. Freedom of speech in many parts of the world, especially the developed countries and liberty of socialization, calls for noteworthy information sharing during the panic pandemic. However, as a communication intervention during crises in the past, social media use is remarkable; the Tweets generated via Twitter during the ongoing COVID-19 is incomparable with the former records. This study examines social media news trends and compares the Tweets on COVID-19 as a corpus from Twitter. By deploying Natural Language Processing (NLP) methods on tweets, we were able to extract and quantify the similarities between some tweets over time, which means that some people say the same thing about the pandemic while other Twitter users view it differently. The tools we used are Spacy, Networkx, WordCloud, and Re. This study contributes to the social media literature by understanding the similarity and divergence of COVID-19 tweets of the public and health agencies such as the World Health Organization (WHO). The study also sheds more light on the COVID-19 sparse and densely text network and their implications for the policymakers. The study explained the limitations and proposed future studies.

Keywords: Social Media, COVID-19, Twitter, Tweets, News, Trend, Natural Language Processing

Introduction
The meteoric rise of social media news during the ongoing COVID-19 is worthy of advanced research. From the records of extant studies, Facebook and Twitter play a predominant role in social information dissemination (Olaleye, Sanusi, and Salo 2018). Freedom of speech in many parts of the world, especially the developed countries and liberty of socialization, calls for noteworthy information sharing during the panic pandemic. Twitter is a preferred social media platform for this study because its information aggregation is methodical, and its communication privilege is snappy.

There are evolving trends in social media in the past. For instance, political communication through e-electioneering has become a trend in many countries globally, and this development is promoting citizen engagement in e-democracy (Macnamara and Kenning, 2011). The current global events and news stories are the trends in China (Yu, Asur, and Huberman, 2011), while Patton, Eschmann, and Butler (2013) examined gang violence, masculinity, and hip-hop in the society through Twitter due to the Internet banging. As a public relations tool for the citizenry, social media is the trend of many local governments across the United States of America (Graham and Avery, 2013).

Though social media use as a communication intervention during crises in the past is remarkable, the Tweets generated via Twitter during the ongoing COVID-19 is incomparable with the former records. Current literature on social media in the context of COVID-19 focused on data and information management from different perspectives. Cinelli et al. (2020) combined five social media platforms and compared questionable and reliable information regarding COVID-19. The authors discovered different information spreading patterns and rumors amplification.

Timely dissemination of information to the public in Vietnam is the focus of the study of La et al. (2020), and the result of their research shows how the combination of policy response, social media, and science journalism helps the Vietnamese fight the deadly Corona Virus promptly. In the same line, Huynh (2020) discussed the risk perception of COVID-19 and concluded that geographical regions and behaviors exhibited while using social media positively affect the risk perception of COVID-19 in Vietnam. Information sharing during the ongoing pandemic is with mixed feelings, and Depoux et al. (2020) argue that the virality of COVID-19 misleading information is as dangerous as the virus itself, and this proposition is in tandem with Pennycook, McPhetres, Zhang, Lu, and Rand (2020). Likewise, the Editorial of Lin (2020) believes that an increase in fear and stigma may escalate through misinformation and misperception of information.

Still, on social media misinformation, Llewellyn (2020) suggests carefulness with trust and expertise on social media, and the author used the personal experience of panic and COVID-19 misleading message saturation as an early warning for the social media users. An earlier study has focused on social media misinformation and information spreading patterns, but a study on close comparison of the public and World Health Organization (WHO) Tweets are limited.

This study intends to examine the trend of social media news and compares the Tweets on COVID-19 as a corpus from Twitter. This study answers the following research questions to explore the set objectives: 1) What is the difference between the public and WHO information spreading patterns via Tweets? 2) How to strike a balance between sparsely and densely tweets during the ongoing COVID-19? The study employed a quantitative methodology and used Natural Language Processing data analysis technique. The study results show three trendy topics and their interrelationship.

This study starts with an introduction, followed by a literature review, methodology, and data analysis. The next part discusses the result, practitioner implication, and conclusion.

**Literature on Natural Language Processing (NLP)**

This section focuses on literature, which deals with the application of NLP on how social media is used as a means of communication during the COVID-19 pandemic to share ideas and thoughts about the effects of the pandemic in different areas of life, with Twitter being the social media platform widely used during the pandemic.

The authors of the reviewed literature cited their ideas using different NLP methods to review people’s thoughts and opinions about the pandemic and its effects on the masses. They also gave an insightful contribution to how social media used during the pandemic, and some of them gave future direction on how social media can be used to solve issues related to the pandemic. The contributions of the synthesized literature is a mixture of positive and negative contradiction. For example, out of thirteen (13) recent papers reviewed, nine (9) studies did not give future direction based on their results. See Table 1 for details.
<table>
<thead>
<tr>
<th>No</th>
<th>Author(s)</th>
<th>Year</th>
<th>Focus</th>
<th>Methodology</th>
<th>Contribution</th>
<th>Future Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kim</td>
<td>2020</td>
<td>This research incorporates a computer-assisted content review to determine which social grooming variables are reducing Incivility of users of social media when writing or reporting on the COVID-19 situation in South Korea.</td>
<td>Binary logistic regression and semantic network analysis</td>
<td>The study indicates that the scale of the social network is a negative indicator of incivility for Twitter users who have built up broader networks and earned supportive feedback from those who are less likely to use uncivil language. The linguistic choice between users varies depending on the scale of the social network.</td>
<td>It suggested that for future work, Social grooming of larger and more diverse social networks is a significant indicator of a decline in civility. It suggest a hybrid study of text and Visual information or other technical research (e.g. visual knowledge) Sentiment Analysis, popularity prediction, virality prediction) as its methodology.</td>
</tr>
<tr>
<td>2</td>
<td>Abd-Alrazaq et al.</td>
<td>2020</td>
<td>The purpose of this study is to identify the key topics posted by users of Twitter concerning the COVID-19 pandemic</td>
<td>Sentiment analysis</td>
<td>The study established 12 subjects, which were grouped into four key themes: the origin of the virus; its source; its effects. People, countries, and the economy; and ways to reduce the risk of infection. The mean sentiment for 10 was good Subjects and negatives on two subjects (deaths caused by COVID-19 and increased racism).</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Cinelli et al.</td>
<td>2020</td>
<td>This research offers an in-depth study of social trends in a time period where theories and moods in social media linked to</td>
<td>Text analysis (NLP) and Mathematical model (SIR model)</td>
<td>The study suggests that the distribution of information is influenced by the model of interactions enforced by specific social media or / or the particular patterns of interaction between groups of users interested in the subject.</td>
<td>-</td>
</tr>
</tbody>
</table>
Sustainable Economic Development and Advancing Education Excellence in the era of Global Pandemic

<table>
<thead>
<tr>
<th></th>
<th>Authors</th>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Sarker &amp; Gonzalez</td>
<td>2017</td>
<td>The research introduces an unlabelled corpus and a collection of language models to the data science, natural language processing and public health communities. The authors built language models to help the development of data mining tools and methods in this area using the data, which are rich in drug-related chatter.</td>
</tr>
<tr>
<td>5</td>
<td>Al-Garadi et al.</td>
<td>2020</td>
<td>The purpose of this paper is to explain the creation and evaluation of a natural language processing and machine learning framework using Twitter data to automatically detect Tweets containing self-reported symptoms of COVID-19-positive patients. The BERT-based model shows high performance in detecting tweets with non-COVID-19 symptoms, while the classifier shows strong performance in detecting tweets with COVID-19 associated symptoms. The authors propose that current performance can be enhanced in the future by incorporating more training data with a wide variety of positive COVID-19 symptoms.</td>
</tr>
<tr>
<td></td>
<td>Authors</td>
<td>Year</td>
<td>Methodology</td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
<td>------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6</td>
<td>Chen et al.</td>
<td>2020</td>
<td>Latent Dirichlet allocation (LDA) and LIWC2015</td>
</tr>
<tr>
<td>7</td>
<td>Mackey et al.</td>
<td>2020</td>
<td>Biterm topic model (BTM)</td>
</tr>
<tr>
<td>8</td>
<td>Li et al.</td>
<td>2020</td>
<td>BERT model, Correlation analysis, and Emotion trend analysis</td>
</tr>
<tr>
<td></td>
<td>Authors</td>
<td>Year</td>
<td>Description</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>9</td>
<td>Banda et al.</td>
<td>2020</td>
<td>This work presents a large-scale curated dataset of more than 152 million tweets, that daily in relation to COVID-19 chatter produced at the time of writing from 1 January to 4 April.</td>
</tr>
<tr>
<td>1</td>
<td>Massaad, &amp; Cherfa</td>
<td>2020</td>
<td>The research is focused on exploring the nature of social media data correlated with Telehealth and public awareness to strategically simplify and speed up the digital health transition</td>
</tr>
<tr>
<td>1</td>
<td>Guo et al.</td>
<td>2021</td>
<td>The authors present a CORD19STS dataset that contains annotated 13,710 Sentence pairs obtained from the open research dataset COVID-19 (CORD19) challenge to translate their output into a Sen-SCI-CORD19-BERT language model</td>
</tr>
<tr>
<td>1</td>
<td>Karisani &amp; Karisani</td>
<td>2020</td>
<td>The focus is to find out to what extent machine learning models are able to distill data generated by the user.</td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Rashed et al.</td>
<td>2020</td>
<td>The authors proposed a toolbox for NLP relating to SARS-CoV-2 and COVID-19.</td>
</tr>
</tbody>
</table>

**Methodology**

**Collection of Twitter Data**

To study similarities between tweets, we chose Twitter as our target field. Twitter has one of the world’s biggest social network platforms and hosts abundant user-generated posts, closely reflecting the public’s reactions towards this pandemic with low expectancy. By deploying Natural Language Processing (NLP) methods on tweets, we were able to extract and quantify the similarities between some tweets over time, which means that some people say the same thing about the pandemic while some Twitter users view it differently. The tools we used are Spacy, Networkx, WordCloud, and Re.

Twitter is a platform to post tweets, share feelings with people, and respond to others’ tweets. Thus, besides likes/replies/retweets, the content of COVID-19 related tweets was mined to see how similar and meaningful some tweets are: Tweet Value.
Since the first confirmed case of the virus in January 2020, #COVID-19 and other similar tags have been trending on Twitter. With thousands of COVID-19 related tweets collected, we wondered how people on Twitter reacted to such tweets over time. Firstly, we explored some engagement metrics of tweets, such as the number of likes. The figure below shows the average likes/replies/retweets per tweet each day. The tweets included COVID-19 related keyword tweets. The tweets were extracted from a large number of everyday opinions.

We used the Twitter data from an open-source, and it was collected throughout 10-weeks, commencing from January 22, 2020, to April 7, 2020. The tweet was in the English language and other languages and contained six variables: public tweets and WHO tweets. A total of 37,315 tweets were captured within the sample. The study selects WHO tweets because it is one of the leading organizations that advocate for the prevention of the COVID-19 disease, and the public information was selected to balance the opinion of people towards the pandemic.

The data was sufficiently high to conduct meaningful analysis, and because of the non-English words in them, during the study’s data-cleaning phase, some tweets that did not contain the Latin alphabets were discarded. A total of 200 tweets was selected from WHO tweets and 600 tweets from public tweets posted in March and April 2020. The time of the tweets extraction was when the virus was assumed to have spread to almost all the continents.

Finally, the final dataset containing 800 tweets was extracted for analysis after cleaning and considering that the dataset was sufficiently high for a meaningful interpretation. The final variables of the data were used for machine learning analysis.
Experiments and Evaluation

The final dataset of 800 tweets for both WHO and public tweets was qualitatively coded. Figure 1 shows the data preparation process from data preprocessing to data processing. The coding process tests the similarities between tweets and the relevance of a coding framework for Covid-19 tweets. These similarities had already been done in a few papers but on sentiment analysis.

First, spacy NLP analysis tools were used to explore the tweet dataset for part of speech, tokenization, and Lemmatization, as shown in Figure 3.3 below.

<table>
<thead>
<tr>
<th>Tweet</th>
<th>Retweet count</th>
<th>Likes count</th>
<th>Tweet value</th>
<th>Date</th>
<th>User</th>
<th>parsed_doc</th>
<th>comment_tokens</th>
<th>comment_lemma</th>
<th>pos_pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5503</td>
<td>19357</td>
<td>950.17</td>
<td>2020-03-11 13:25:56</td>
<td>Judd Legum</td>
<td>{TRUMP TWO, WEEKS, AGO, ..}, {TRUMP, two, weeks, ago, ..}, {TRUMP, two, weeks, ago, ..}, {TRUMP, two, weeks, ago, ..}, {PRON, have, 15, ..}, {PRON, have, 15, ..}, {PRON, have, 15, ..}, {PRON, have, 15, ..}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1603</td>
<td>6231</td>
<td>2729.58</td>
<td>2020-03-11 13:18:53</td>
<td>Diamond &amp; Silk</td>
<td>{it, looks, like, the, Democrats, .., along, at, ..}</td>
<td>{it, looks, like, the, Democrats, .., along, at, ..}</td>
<td>{it, looks, like, the, Democrats, .., along, at, ..}</td>
<td>{it, looks, like, the, Democrats, .., along, at, ..}</td>
</tr>
<tr>
<td>2</td>
<td>1510</td>
<td>5032</td>
<td>12.47</td>
<td>2020-03-11 13:17:13</td>
<td>Sebastian Campos C.</td>
<td>{Gente, .., con, todo, esto, del, coronavirus, ..}</td>
<td>{Gente, .., con, todo, esto, del, coronavirus, ..}</td>
<td>{Gente, .., con, todo, esto, del, coronavirus, ..}</td>
<td>{Gente, .., con, todo, esto, del, coronavirus, ..}</td>
</tr>
<tr>
<td>3</td>
<td>4924</td>
<td>6380</td>
<td>435.53</td>
<td>2020-03-11 13:16:45</td>
<td>David Gay Johnson</td>
<td>{Brooklyn, teacher, .., returning, from, italy, ..}</td>
<td>{Brooklyn, teacher, .., returning, from, italy, ..}</td>
<td>{Brooklyn, teacher, .., returning, from, italy, ..}</td>
<td>{Brooklyn, teacher, .., returning, from, italy, ..}</td>
</tr>
<tr>
<td>4</td>
<td>29346</td>
<td>188955</td>
<td>9.02</td>
<td>2020-03-11 13:14:08</td>
<td>Georgiak</td>
<td>{italy, .., on, lockdown, bc, of, coronavirus, ..}</td>
<td>{italy, .., on, lockdown, bc, of, coronavirus, ..}</td>
<td>{italy, .., on, lockdown, bc, of, coronavirus, ..}</td>
<td>{italy, .., on, lockdown, bc, of, coronavirus, ..}</td>
</tr>
<tr>
<td>5</td>
<td>3215</td>
<td>17929</td>
<td>22.20</td>
<td>2020-03-11 13:09:43</td>
<td>Lenziska Lima 60</td>
<td>{Cuba, anunció, vacina, contra, o, Coronavirus}</td>
<td>{Cuba, anunció, vacina, contra, o, Coronavirus}</td>
<td>{Cuba, anunció, vacina, contra, o, Coronavirus}</td>
<td>{Cuba, anunció, vacina, contra, o, Coronavirus}</td>
</tr>
</tbody>
</table>

Fig 3.3: Glance at the tweet document and parsed documents

Furthermore, we analyze the similarities in the tweets on a more detailed level, comparing two rows in the tweet data to find similarities. To further clean the dataset, stopwords are removed from the dataset. Stopwords are commonly and frequently used words that search engines or analysis packages are programmed to ignore when searching and retrieving results of a search query. Examples of stopwords include elsewhere, against, on, and alike (Geeks for Geeks, 2016).

We conducted further analysis by utilizing the spacy library. Spacy is a free, open-source library for Natural Language Processing in Python. It features NER, Part of Speech (POS) tagging, dependency parsing, word vectors, and many more (Spacy, 2020). Here in this analysis, we used it in combination with NetworkX to form nodes and edges linking similar words together, basically forming a graph. NetworkX library is used for studying graphs and networks.

Here is how it was generated: The first step was to prepare the spacy library dataset to analyze from, where we find similarities between two tweets. However, in similarities between two tweets, sentence embedding is constructed on spacy by averaging the word embeddings and, most importantly, meaningless words such as stopwords in ordinary sentences. We use the Word2Vec model to calculate the sentence similarity in tweets since the tweet sentences are composed of words. We created another column of parsed_doc to parse the tweets to its component parts, as shown in fig 3.3. For the second step, we introduced the NetworkX to form graphs to link similar words, as shown in Fig 4.4.

In the third step, we defined some metrics to help further comprehend the changes in our analysis. We first tokenize the tweets to separate them into smaller units, which is also included in the spacy library. After this separation, we introduced the Word Cloud package to visualize the tweet dataset. Word Cloud library in python is a data visualization technique used for representing text data in which the size of each word indicates how often it appears or its importance in a sentence or list.

Findings

After putting our model results in the timeline under the pandemic context, we summarized some interesting findings.
Sustainable Economic Development and Advancing Education Excellence in the era of Global Pandemic

'the most effective way to wash your hands stay clean my friends #coronavirus #COVID https://t.co/K1rkco6yNg'

data.Tweet[14]
'these are 7 simple steps to protect yourself and others from #COVID19.\nhttps://t.co/RU2zgr65jm https://t.co/TNvSy9P0u'

data.Tweet[17]
'every single person should be participating in Coronavirus response by taking the same simple precautions every day.\nmore from Ambassador Brrn: https://t.co/8gElkygTiK'

data.Tweet[13]
'Nottingham Forest confirm all players, staff and officers have been tested for the coronavirus and all results have come back negative.'

data.Tweet[15]
'one of the first Americans to be diagnosed with Coronavirus (a 48-year-old man from Rhode Island) shares his experience with W S3\n\n\nIt almost killed me. Its alarming when I hear people minimize it as a cold. It was no simple cold for me. \n\nhttps://t.co/fAvsoMKQ4Q https://t.co/qmEB22b4A'

data.Tweet[13]
'Nottingham Forest confirm all players, staff and officers have been tested for the coronavirus and all results have come back negative.'

data.Tweet[14]
'these are 7 simple steps to protect yourself and others from #COVID19.\nhttps://t.co/RU2zgr65jm https://t.co/TNvSy9P0u'

data.Tweet[15]
'one of the first Americans to be diagnosed with Coronavirus (a 48-year-old man from Rhode Island) shares his experience with W S3\n\n\nIt almost killed me. Its alarming when I hear people minimize it as a cold. It was no simple cold for me. \n\nhttps://t.co/fAvsoMKQ4Q https://t.co/qmEB22b4A'

data.Tweet[17]
'every single person should be participating in Coronavirus response by taking the same simple precautions every day.\nmore from Ambassador Brrn: https://t.co/8gElkygTiK'

data.Tweet[40]
'knowing the symptoms of #COVID19 will help you get the treatment if you need it. Learn how to tell the difference between #Coronavirus #flu and the #CommonCold.\n
\n\nspread the message not the virus.\n\nHSElive | @HSEIreland | @Bscireland https://t.co/mBk7sW6m3'

data.Tweet[61]
'Coronavirus conference canceled in New York because of coronavirus https://t.co/qorEtBkUm https://t.co/OMw4laL0vb8'
As shown in the figure above, there is a similarity between the tweet in row 11 and 14. This comparison accounts for 89.57% of the Twitter dataset. Likewise, tweets 14 and 17 with a similarity value of 84.60%, and tweet 13 and 15 give a similarity value of 86.95%. The similarity values are strong and conform to the threshold of > 0.7 (Gerstman, 2008; 2014). However, even though the similarity is high between these tweets because of the messages they carry, it does not necessarily mean it valued the same; that is, the importance allocated to them because of the user's opinion, as shown in figure 4.2. We also checked the likeness between tweets 40 and 61, which gave a similarity value of 68.65%. This similarity value is moderate and conforms to the boundary of 0.7 (Gerstman, 2008; 2014). Their similarity is relatively low compared to the other public tweets' similarities (see Figure 4.1).
Likewise, there are similarities between public and WHO tweets, but not similar to having the tweets from the same source, as seen between public tweets. As indicated in figure 4.3, tweets 14 and 89, tweets 11 and 12, 4, 112, and 61 give a similarity value of 74.62%, 74.68%, and 69.74%. Tweets 13 and 95 shows a similarity value of 49.50%. The first and second values conform to the threshold value > 0.7 while the third and fourth similarity values conform to moderate (see figure 4.1).

Furthermore, we linked similar words together with networkX, as mentioned earlier. To make the graph legible enough, we represented 70 tweets on the networkX graph, and we showed just a part of the graph in the figure below:

**Fig 4.3: Similarity results between public tweets and WHO tweet**

With the topic modeling, we extracted some hidden topics summary from the tweet, and each topic is considered a collection of keywords in some proportion. Extracting a good quality topic depends on the text used in the analysis. We used Latent Dirichlet Allocation (LDA) from the Gensim package (a very efficient LDA implementation). After
filtering tweets by keywords, suggested by the frequency of some words and the topic trends, we modelled ten topics, as shown in fig 4.4.

The first thing we did was to get the top-30 most salient terms. Saliency is a measure of how much the term tells us about a topic, while Relevance is the weighted average of the word's probability, given the topic normalized by the probability of the topic. The bubble in the inter-topic distance map measures the importance of the topics relative to the data. The larger the bubble, the more critical the topics are. As shown in fig 4.4, specific topics clustered together, indicating a similarity between topics, and the more the topics, the more clusters form and similarities.

The same trend of news topics applies here as tweets mentioning CORONAVIRUS, and some related words to its outbreak spread from China to other parts of the world. As shown in the table and word cloud, the public paid more and more attention to government actions over time. Medical-related, economic impact, and life during COVID topics increased slowly. The game season, mask, and stay at home topics did not show much upward trend over time with the tweet from the period taken.

Table 4.1: Table showing 5 topic model for public tweet

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 1 weight</th>
<th>Topic 2</th>
<th>Topic 2 weight</th>
<th>Topic 3</th>
<th>Topic 3 weight</th>
<th>Topic 4</th>
<th>Topic 4 weight</th>
<th>Topic 5</th>
<th>Topic 5 weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronavirus</td>
<td>0.059</td>
<td>coronavirus</td>
<td>0.033</td>
<td>coronavirus</td>
<td>0.092</td>
<td>coronavirus</td>
<td>0.055</td>
<td>confirm</td>
<td>0.040</td>
</tr>
<tr>
<td>Spread</td>
<td>0.038</td>
<td>outbreak</td>
<td>0.031</td>
<td>cancel</td>
<td>0.044</td>
<td>people</td>
<td>0.035</td>
<td>coronaviru s</td>
<td>0.035</td>
</tr>
<tr>
<td>stop</td>
<td>0.030</td>
<td>need</td>
<td>0.018</td>
<td>day</td>
<td>0.020</td>
<td>virus</td>
<td>0.022</td>
<td>break</td>
<td>0.032</td>
</tr>
<tr>
<td>amp</td>
<td>0.021</td>
<td>medium</td>
<td>0.017</td>
<td>response</td>
<td>0.019</td>
<td>test</td>
<td>0.022</td>
<td>online</td>
<td>0.018</td>
</tr>
<tr>
<td>patient</td>
<td>0.020</td>
<td>disinfect</td>
<td>0.013</td>
<td>rate</td>
<td>0.018</td>
<td>think</td>
<td>0.020</td>
<td>report</td>
<td>0.016</td>
</tr>
<tr>
<td>learn</td>
<td>0.020</td>
<td>space</td>
<td>0.013</td>
<td>get</td>
<td>0.017</td>
<td>know</td>
<td>0.019</td>
<td>email</td>
<td>0.016</td>
</tr>
</tbody>
</table>
When visualizing using word clouds, we can see the high importance given to the word CORONAVIRUS so many times as well as some words like a test, flu, people, break, virus, public, china, cases, spread, death, hospital, trump, response, and new cases seems to be some of the most used words in the tweets from January 22, 2020, to April 7, 2020.

Practitioner Implications

One of the recent studies established the trust level of online news, mainly social media and search engines, and concludes that trust levels on traditional news media are higher than social media (Park, Fisher, Flew, and Dulleck, 2020). In like manner, Llewellyn (2020) raised some warning signals as an antecedent of social media message trust. Besides, other authors have raised their concern on rumors amplification, information dissemination timeliness, location, behaviour, fear, and stigma as the source of social media information misinformation, misconception, and misperception (Cinelli et al. 2020; La et al. 2020; Huynh (2020; Depoux et al. 2020; Lin 2020; McPhetres, Zhang, Lu, and Rand 2020). With arguments and proposition from the earlier studies, it became apparent that multiple
uncensored sources of information through social media is a challenge that needs quick intervention. Social media are confused about whom to trust, public or government. If it is possible to tread the line of authority and specialization, WHO information should be more reliable for the public and preferable to the public information. This study answers the research question of the difference between public and WHO information spreading patterns via Tweets (RQ1).

The study shows a close similarity of tweets between the public and the WHO, but in terms of the worth and weight, WHO’s tweet value, surpasses the public tweet. To go by this result, people still perceived the WHO information dissemination through tweets to be reliable and dependable despite the bias and sentiment of misinformation. Another question that prompts this finding is why WHO information on traditional media has the same equal weight with social media through the same content but a different channel. The earlier studies mentioned above show an imbalance of social media messages.

Fig 4.8: Topic Modelling graphs for WHO tweets

The study shows a close similarity of tweets between the public and the WHO, but in terms of the worth and weight, WHO’s tweet value, surpasses the public tweet. To go by this result, people still perceived the WHO information dissemination through tweets to be reliable and dependable despite the bias and sentiment of misinformation. Another question that prompts this finding is why WHO information on traditional media has the same equal weight with social media through the same content but a different channel. The earlier studies mentioned above show an imbalance of social media messages.
Table 4.2: Table showing 5 topic model for WHO tweet

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
<th>Topic 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>person</td>
<td>right</td>
<td>people</td>
<td>free</td>
<td>0.028</td>
</tr>
<tr>
<td>health</td>
<td>coronaviruses</td>
<td>need</td>
<td>ill</td>
<td>coronavirus</td>
<td>0.029</td>
</tr>
<tr>
<td>system</td>
<td>use</td>
<td>people</td>
<td>vulnerable</td>
<td>maintain</td>
<td>0.022</td>
</tr>
<tr>
<td>vital</td>
<td>home</td>
<td>prepare</td>
<td>household</td>
<td>course</td>
<td>0.019</td>
</tr>
<tr>
<td>patient</td>
<td>confirm</td>
<td>service</td>
<td>amp</td>
<td>language</td>
<td>0.019</td>
</tr>
<tr>
<td>shortage</td>
<td>suspect</td>
<td>world</td>
<td>coronavirus</td>
<td>product</td>
<td>0.019</td>
</tr>
<tr>
<td>coronavirus</td>
<td>mask</td>
<td>coronavirus</td>
<td>country</td>
<td>work</td>
<td>0.018</td>
</tr>
<tr>
<td>mask</td>
<td>refugee</td>
<td>global</td>
<td>work</td>
<td>join</td>
<td>0.014</td>
</tr>
<tr>
<td>incl</td>
<td>people</td>
<td>support</td>
<td>water</td>
<td>area</td>
<td>0.012</td>
</tr>
<tr>
<td>improve</td>
<td>cough</td>
<td>government</td>
<td>soap</td>
<td>demand</td>
<td>0.012</td>
</tr>
</tbody>
</table>

It is a combination of fake and genuine news with mixed reactions. The study also answers the question of how to strike a balance between sparsely and densely tweets during the ongoing COVID-19 (RQ2). This study established COVID-19 sparsely and densely tweets. 

![Intertopic Distance Map (via multidimensional scaling)](image)

![Top-30 Most Salient Terms](image)

Fig 4.9: Topic Modelling graphs for Public and WHO tweets
This comparison indicates that the similar tweets clustered together and densely populated while the unfamiliar tweets lightly clustered together and sparsely recognized and indicate different opinions. The dense tweets show that COVID-19 tweets of highly connected opinions are more equally linked together, which might be advantageous for policymakers' information processing. The revealed dense networks can give COVID-19 managers new ideas about the COVID-19 message's sentiments and how COVID-19 spread in an increasingly interconnected society, and how this affects the increasingly connected society. This insight helps the manager's position their strategy on transferring information in a very transparent manner and positively influencing the societal decision. This study recommends that the managers will strike a balance between sparsely and densely tweets when leveraging on it for decision making.

**Conclusion**

In addition to what earlier studies emphasized, social media's importance and danger in information dissemination is escalating, especially during the ongoing COVID-19. This study contributes to the social media literature by understanding the similarity and divergence of COVID-19 tweets of the public and the health agency such as WHO. The study also sheds more light on the COVID-19 sparse and densely text network and their implications for the policymakers. Despite the contribution of this study, it is limited by time-series data of COVID-19. As COVID-19 prediction is curvilinear, its dataset is changing. Future studies should collect more COVID-19 tweet data and embark on a comparison study to give meaning to the past and current data.

**References**


5973


\textbullet{} La, V. P., Pham, T. H., Ho, M. T., Nguyen, M. H., P Nguyen, K. L., Vuong, T. T. & Vuong, Q. H. (2020). Policy response, social media and science journalism for the sustainability of the public health system amid the COVID-19 outbreak: The vietnam lessons. \textit{Sustainability}, 12(7), 2931.


