What people share about the COVID-19 outbreak on Twitter? An exploratory analysis

Dhivya Karmegam, Bagavandas Mapillairaju

ABSTRACT
Background The recent outbreak of respiratory illness caused by COVID-19 in Wuhan, China, has received global attention as it has infected thousands of individuals there, and later it has also been reported from other countries internationally. This study aims at performing an exploratory study on Twitter to understand the information shared among the community regarding the COVID-19 outbreak.

Methods COVID-19 related tweets were collected from Twitter using keywords from 18 January to 25 January 2020. Top-ranking tweets were taken as samples and then categorised based on the content. Expressions or opinion tweets were analysed qualitatively to understand the mindset of the people regarding the outbreak. Theme wise reachability evaluation of the messages was also performed.

Results Based on the content of the tweets, five themes were evolved: (1) general information; (2) health information; (3) expressions; (4) humour and (5) others. 57.42% of messages are general information followed by expressive tweets (24.12%). Humorous messages were liked the most, whereas health information tweets were retweeted the maximum. Fear was the predominant emotion expressed in the messages.

Conclusion The results of the study would be useful to focus on the dissemination of the right information and effective communication on Twitter related to health and outbreak management.

INTRODUCTION
In recent days, COVID-19 has conquered the headlines in the news media and captured global attention as it infected more than thousands in China, and also it has been spread to other countries globally. COVID-19 (also called 2019-nCoV) causes severe respiratory illness, and it was first reported and detected in Wuhan City, China, in December 2019. As of 26 January 2020, COVID-19 infections were also reported from many other countries that include USA, Australia, Japan, France and so on. Controlling and preventing the outbreak need continuous monitoring and international collaboration. There is a need for quick access to information regarding the outbreak, and the public should be aware of the risks and challenges involved in managing and preventing the COVID-19 infection. At the beginning of the outbreak, quick communication of information to the public is vital to prevent the fast-spreading virus. Communication regarding health during the outbreak will be effective only if it is done from the perspective of the people and also the content should be relevant to the targeted people. Understanding the perception of the people regarding the virus infection and outbreak by traditional data collection techniques has a lot of practical difficulties. It consumes a lot of time and money. When there is a crisis, people tend to use social media to seek information and express their views and feelings. In the case of COVID-19 outbreak, when physical distancing, isolation and travel restrictions are imposed, social media turns out to be the forum for public discussion more than usual. Messages shared by the people in social media like Twitter can be used to understand the expectations of the people and their mindset. In turn, this will help in planning, communicating and managing the crisis effectively based on the needs of the public. Social media has the potential to assess the awareness level and the state of mind of the public regarding any particular unexpected event. Previous studies also confirm that such analysis helps the public health stakeholders to know about information dissemination immediately during such an infectious disease outbreak. The earlier researches confirm that social media is one of the effective tools used to communicate information regarding preventive measures during the outbreak. The analysis of the content of the messages in social media provides an insight regarding the awareness and the needs of the public. There is a gap between the information spread over by the media and the concerns of the people. The purpose of this study is to understand the information and...
content communicated on Twitter from 18 January to 25 January 2020 by the people globally about the COVID-19 outbreak. This helps the public health stakeholders to plan Twitter communication and information dissemination regarding the outbreak and associated information. To use social media to disseminate the right information regarding COVID-19 pandemic to reach a wider public, the initial and most important step is to understand the content of the messages posted and shared by the people. The objectives of this study are:
1. To categorise the Twitter messages related to COVID-19 infection based on their content into themes.
2. To explore the opinions and emotions of the people regarding the COVID-19 outbreak qualitatively on Twitter.

METHODS
Twitter messages were collected from Twitter search Application Programming Interface using hashtags related to the COVID-19 outbreak for 7 days from 18 January to 25 January 2020. Original tweets (excluding retweets) in English were collected using the following hashtags: coronavirus, wuhan, coronaviruswuhan, coronaoutbreak, wuhanvirus, WuhanCoronavirus, WuhanPneumonia, ChinaCoronaVirus, ChinaWuHan, 2019nCoV, WuhanOutbreak, VirusCorona, nCoV2019 and nCoV. The collection of tweets using these keywords was done using R programming, and it was 100,500 tweets from 61,216 unique users. Among the tweets collected, only 2,409 tweets were georeferenced. During the considered period, maximum numbers of confirmed cases of COVID-19 were reported from the countries that do not use Twitter or from non-native English-speaking countries. These aforementioned limitations restricted us to further analyse the tweets based on their geo-location.

Sample tweets were extracted from collected tweets by top-ranked sampling strategy. The favourite count (number of times the message was liked) and the retweet count (number of times the message was shared) of a particular tweet is the indicator of reachability of the tweets. The higher the favourite or the retweet count greater than 100 would be considered as top-ranked tweets, and they were taken as samples (n=1219).

In order to understand the details and information shared by the people about outbreak in Twitter, content analysis was performed. Based on the content of the messages, themes (what is discussed in the tweet) were identified and coded by the direct content analysis method. The first author framed a codebook of themes with reference to the earlier studies and a general look at the messages. Based on the codebook, both the authors coded 100 randomly selected tweets independently. Inconsistencies in coding between the two authors were discussed and resolved while finalising the codebook of themes. With reference to the finalised codebook, both the authors (DK and BM) assigned the themes to the sample tweets independently. All the differences in coding were discussed by both the authors together until an agreement in final coding was met. Twitter messages that contained more than one theme were assigned as such. Then, the theme-wise reachability of the messages was evaluated based on the favourite and retweet ratio as indicators. The retweeting frequency and number of times the tweets were liked (favourite count) were used as the measure of interactivity and reachability of the tweet. The retweet and favourite rate were calculated as a ratio by dividing the total retweet or favourite count corresponding to a particular theme by the total number of messages on that theme.

To understand the mindset of the people regarding the COVID-19 outbreak and to explore the emotions expressed regarding the situation, tweets that stated expressions and opinions were examined qualitatively by discourse analysis. As Twitter messages had a limited number of words with hidden information, discourse analysis was done. Expressive and opinion tweets were read many times by both the authors, and the categories that evolved from the text were identified. The categories identified and the contexts behind the expressions were discussed qualitatively with example tweets.

RESULTS
Content analysis
Five themes were identified from the randomly selected 100 tweets. They were health information, general information and updates, expressions or opinions, sarcasm or humour, and others. Personal conversations and tweets that were not related to COVID-19 outbreak were categorised as a theme, others. Out of 1219 sample records, 98 (8.04%) tweets were coded under others category. The maximum number of messages contains general information and updates followed by expression or opinion tweets were many. Table 1 describes the themes with example tweets and the number of tweets in each category.

Online supplemental file 1 gives the status ID of the sample tweets, corresponding theme, favourite count and retweet count. Each tweet had a maximum of two themes.

Among the four major themes, sarcasm or humour had the lowest count but the highest favourite ratio (925). The sarcastic tweet’s favourite ratio of 925 indicates that each sarcastic message was likely to be favourited 925 times on an average. Tweets from health information had the highest retweet ratio (577). The higher the retweet ratio, the higher will be the shares and reachability of that content to the Twitter users. Table 2 shows the theme-wise favourite count, retweet count, and favourite and retweet ratio.

Qualitative analysis
The tweets that stated opinions and expressions were analysed qualitatively to understand the mindset,
emotions and the context behind the emotional expression. The subcategories that emerged from opinion or expressive tweets were negative emotions, positive and thankful, suggestions and seeking information. As the results were presented qualitatively, the frequency of the tweets under each sub-category was specified using terms such as ‘often’, ‘frequent’ and ‘some’.\textsuperscript{30, 31} The opinions, emotions and the context of expressions were explained qualitatively under each subcategory with example tweets.

**Negative emotions**

People expressed negative emotions such as fear, anger, sadness and frustration towards the outbreak situation. Fear was frequently revealed in the messages. Fear was expressed by the people on knowing the statistics on confirmed cases and deaths, transmission routes and also on watching live videos of hospitals. Some example tweets that indicated fear were given below.

The Videos Of People Collapsing On Chinese Streets Due To The Coronavirus Are Scary As Shit.

This is very bad news, meaning airports checkpoints don’t work since some #coronavirus patients are asymptomatic and will continue spreading 2019-nCoV without even knowing. Nightmare scenario!

SCARY! South China Morning Post reports that 41 people are DEAD from coronavirus in China with more than 1000 people are infected.

Followed by fear, sadness was sensed often in the tweets. Sadness was expressed concerning the outbreak circumstances, affected people and deaths. Few sample tweets that show sadness were presented.

i swear it was only last week it had breached no national boundaries. now it’s breached 7? that seems like a pretty rapid acceleration, how worried are we supposed to be.

My heart breaks for the poor people dying of the Coronavirus. They’re finding them all over.

**Table 1** Themes with description, example tweets and the number of tweets in each category

<table>
<thead>
<tr>
<th>Themes</th>
<th>Description</th>
<th>Example tweets</th>
<th>Number of tweets (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health information</td>
<td>Tweets that had information regarding transmission, symptoms, prevention and treatment for respiratory illness caused by COVID-19.</td>
<td>#'WuhanCoronavirus Symptoms of 2019-nCoV include: A fever Dry cough Shortness of breath Trouble breathing Dear all: frequent hand washing, covering coughs, and avoid touching your face <a href="https://t.co/jdhHu203Yr">https://t.co/jdhHu203Yr</a>'.</td>
<td>138 (11.32)</td>
</tr>
<tr>
<td>General information and updates</td>
<td>Tweets that give updates on the count of confirmed cases and mortality in different countries. Tweets that share information regarding the steps carried out by the government in airports and hospitals to control and prevent outbreaks. Videos and images of the situation at hospitals in China.</td>
<td>'China has locked down a city of 11 million to slow the spread of the new strain of coronavirus. Latest estimates: ► 17 dead ► 4k infected ► 22+ cities affected in 5–8 countries ► WHO will soon decide on a declaration of global public health emergency'.</td>
<td>700 (57.42)</td>
</tr>
<tr>
<td>Expressions or opinions</td>
<td>Messages in which people expressed their opinions, emotions and clarifications regarding the outbreak.</td>
<td>'What a nightmare. There is no way any city this size would be comfortable in lockdown'.</td>
<td>294 (24.12)</td>
</tr>
<tr>
<td>Sarcasm or humour</td>
<td>Humorous or witty text, sarcastic images, videos and memes related to the COVID-19 outbreak.</td>
<td>'a little late … The coronavirus has already gone into &quot;tourist mode&quot;, currently taking airline trips around the globe'.</td>
<td>92 (7.55)</td>
</tr>
</tbody>
</table>

**Table 2** Theme-wise favourite count, retweet count, and favourite and retweet ratio

<table>
<thead>
<tr>
<th>Themes</th>
<th>Favourite count</th>
<th>Retweet count</th>
<th>Favourite ratio</th>
<th>Retweet ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health information</td>
<td>89 585</td>
<td>79 578</td>
<td>649</td>
<td>577</td>
</tr>
<tr>
<td>General information and updates</td>
<td>382 474</td>
<td>258 619</td>
<td>546</td>
<td>369</td>
</tr>
<tr>
<td>Expressions or opinions</td>
<td>259 600</td>
<td>87 287</td>
<td>883</td>
<td>297</td>
</tr>
<tr>
<td>Sarcasm or humour</td>
<td>85 098</td>
<td>24 851</td>
<td>925</td>
<td>270</td>
</tr>
</tbody>
</table>

Frustration, anger and concerns were also exhibited in some of the messages along with other emotions in the context of outbreak settings. In some tweets, people stated that there was a delay in the actions taken by the government in controlling the spread. Some tweets also disclosed the lack of trust in the initiatives taken by the government and other authorised institutions to control the outbreak. Discontent regarding the rumours and misinformation spread in the social media was also voiced in some of the tweets. Some of the sample tweets that expressed anger and frustration were given.

this coronavirus shit is nuts.
there’s a new f*ckin disease that’s already killing people and is confirmed to be in the United States.
LIES!!! Why are we allowing people from Wuhan in Toronto after the quarantine was announced? This is irresponsible!!
let’s just spread a bunch of disinfo about the coronavirus who-fucking cares anymore.

Positive and thankful
Positive opinions and supportive thoughts were shared in some Twitter messages towards the action taken by the officials and stakeholders while handling the crisis. People also expressed their gratitude to the public health professionals for their patience and their concern.

#CoronavirusOutbreak in #Wuhan goes viral and touches many; respects and tributes go to all of those doctors, nurses, and scientists.
What wasn’t inevitable: screening on presentation to the ER, immediate isolation, calling @TOPublicHealth. The fact that this occurred shows the system is working. Kudos to staff at @Sunnybrook for your diligence.

Suggestions
In some messages, the public stated their personal opinions and suggestions regarding what could be done to control the virus spread and outbreak. One of the frequent suggestions mentioned in messages was closing the international borders thereby restricting people from other countries. Sample opinion tweets were given below.

Now it’s spreading in the United States. Two words, TRAVEL BAN.
With Coronavirus Spreading, Canada Must Restrict Incoming Flights From China.

Seeking information
People also look for information and asks for clarification in relation to the outbreak. The public seeks information regarding the prevention and treatment, updates on affected cases and government initiatives. Twitter messages in which people asked questions regarding the outbreak were provided.

So when does the WHO declare coronavirus an emergency?
Anyone have a total confirmed coronavirus case number?
Is sushi safe to eat with this coronavirus shit going on?

DISCUSSION
The main aim of this study was to understand the content of the Twitter messages regarding the COVID-19 outbreak. From the results, it was clear that a maximum of general information and updates regarding the virus and outbreak was shared in social media. The next highest content in the Twitter messages was personal opinions or expressions shared by the people. This implies that people were aware of and also concerned about the spread of COVID-19 infections. Although the considered period in this study was the initial stage of the pandemic, we see a significant count of health information posted and shared. Health information messages posted would have provided the public, the knowledge regarding the symptoms and the prevention of COVID-19. Based on the reachability evaluation, even though sarcastic or humorous messages were liked the most, health information messages were likely to be retweeted and shared the maximum. This may be because people would have thought sharing health information was vital so that it reaches many Twitter users. However, regardless of the situation, people enjoyed the humorous tweets and memes posted.

Based on our qualitative analysis, we see that the public expressed mixed emotions and opinions (both positive and negative). Fear and sadness were predominantly prevailing among the people. The live videos of people suffering in hospitals, images of crowded hospitals, ambulances waiting with patients outside the hospital, incorrect personal opinions and information with more hyped numbers of the infected and mortality than the actual were widely shared in the Twitter messages. Some articles shared on Twitter presented the COVID-19 as a deadly and killer virus. These videos, images and personal opinions may be the reason for panic and unrest among the people. The information seeking behaviour of the public approves the responsiveness and concern of the public towards the crisis.

The evidence of misinformation in social media was recognised widely. Also the users did not share the link from the authorised website and sources directly on social media. Hence, it is vital to check the credibility of the messages to handle the situation efficiently. Studies say that unverified content and rumours were shared mostly by individual accounts when compared with other sources like non-governmental organisations, media and healthcare accounts. As per the experts, one of the best approaches to avoid misinformation in social media is to increase social media messages from the authorised and official sources. Healthcare professionals, researchers and famous personalities can be used
as a source to load social media with trustworthy information. This will increase the verified information in social media, so that information and opinions of individuals may not dominate the social media content. If the information shared by the official sources satisfies the expectations and concerns of the people, then there will be no space for misinformation becoming viral.

Themes and expressions extracted from the tweets in this exploratory analysis give a basic understanding of the content, people’s expectations and concerns. Based on this, outbreak managers can strategise communication by sharing messages that respond to public concerns and feelings especially fear. Clear, short, practical and shareable messages in social media can easily reach the audience. Other features of communication like multimedia and videos with illustrated comics may also help in spreading the information faster. Although the potential of social media analysis in crisis management and communication through the media has widely drawn the attention of the researchers, it has never been integrated into practice for outbreak management. The public health organisations and stakeholders need to understand the content, communicating features and trends in the social media and take advantage of those features to strategise the right method of information dissemination to the wider audience. An analysis of content and expressions were provided in this study and that can be used while planning communication of verified information dissemination.

This study analysed the data collected in 1-week duration. Also only a sample of tweets was considered for analysis. So, there is a possibility of limitations in the results arrived at using the limited data. One source (Twitter) of information of public views was examined in this study. Furthermore, the tweets only in English were considered. So there is a chance of losing the information from the messages posted in other social networking sites and other languages. However, this initial effort of content analysis of outbreak-related tweets gives a basic understanding of the situation and helps in planning the information dissemination on Twitter and other social media during the outbreak.

CONCLUSION

The study gives an overview of what has been shared on Twitter regarding the outbreak. Looking at the crisis from the point of view of the people and understanding their mindset is vital for the stakeholders in planning the response. COVID-19 related tweets were rich sources of information and opinions. As more information about the virus and the outbreak were posted and shared on Twitter, the same can be used by the public health professionals to disseminate the right information at the time of crisis. The results arrived at after examining the expression tweets give a basic understanding of people’s needs and their state of mind towards the outbreak. Instead of letting the different opinions of the people dominating the content of social media, health communications on Twitter can be strategised in such a way that it was replaced by scientific knowledge. Precise and verified information about virus transmission and prevention, communicated at the right time to the people, will improve their behavioural response and also help in reducing fear. This study analysed the tweets at a global level, whereas examining the tweets at the national level integrated with network analysis will give more detailed information for decision making.

REFERENCES


ORCID iD

Dhivya Karmegam http://orcid.org/0000-0003-3307-8704
16 Carter M. How Twitter may have helped Nigeria contain Ebola. BMJ 2014;349:g6946.
36 Llewellyn S. Covid-19: how to be careful with trust and expertise on social media. BMJ 2020;368:m1160.