

An exploratory analysis of public opinion and sentiments towards COVID-19 pandemic using Twitter data.

Emeka Chukwusa (✉ emeka.chukwusa@kcl.ac.uk)

Kings College London

Halle Johnson

Kings College London

Wei Gao

Kings College London

Research Article

Keywords: Coronavirus, Social media data, Twitter, Sentiment analysis and COVID-19 Pandemic.

DOI: <https://doi.org/10.21203/rs.3.rs-33616/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective: Twitter data have been increasingly used to address health-related issues. However, little is known about their potential for understanding public opinions and sentiments of the current COVID-19 pandemic. The present study explores public opinion and sentiments about the COVID-19 pandemic using Tweets from 3 popular Coronavirus-related hashtags (#COVID19, #Coronavirus, #SARSCoV2).

Results: Of the 39,726 Tweets analysed, we found that over 60% of words used within Tweets in all hashtags (#COVID19, 63.9%; #Coronavirus: 65.6%; #SARSCoV2: 63.5%) conveyed a negative mood towards the pandemic. Our results also showed similar trends in Tweet volume in #COVID19 and #SARSCoV2, with a spike in the number of Tweets on the 3rd and 6th of April 2020. Further exploration of Tweets in both hashtags revealed similar Twitter discussions related to topics on “Hydroxychloroquine” and “Hospitalisations of the British Prime minister” and “ the attainment of 1 million cases of coronavirus globally”.

The findings of this exploratory study indicate that there is potential for using data generated from Twitter to understand general public opinion and sentiments towards the COVID-19 pandemic. However, caution is needed due to several limitations in this study. It is also important for future studies to explore the context around Tweets.

Introduction

The COVID-19 pandemic, currently affecting millions of people globally, has resulted in a rise of the use of social media platforms (e.g. Twitter, Facebook, Instagram, Weibo, TikTok and WeChat). Due to social distancing and movement restrictions, many individuals and organisations are taking to social media to communicate, share and express their opinions, sentiments and experiences about the pandemic.

Twitter is one of the most popular social media platforms. It is a microblogging service with over 300 million active users [1]. Twitter users share personal opinions and experiences by posting 280-character messages known as Tweets. Twitter has been increasingly used as a data source for health-related research, offering a more efficient means of data collection over traditional survey methods [2]. For instance, it has been used to examine disease stigma [3], and to monitor disease pandemics [4] [5]. More recent studies have used Twitter data for the assessment of public sentiments, attitudes and opinions about health related issues [6] [7]. These studies have shown that there may be scope for using Twitter generated data to provide insight about the public opinion on the current COVID-19 pandemic.

The present study aims to explore the use of Twitter data to understand public opinion and sentiment about the COVID-19 pandemic using Tweets from three popular coronavirus-related hashtags (#COVID19, #Coronavirus and #SARSCoV2).

Methods

We downloaded publicly available Tweets from three popular Coronavirus-related hashtags (#COVID19; #SARCoV2 and #Coronavirus). Tweets were consecutively downloaded over a 7-day period (1st April - 7th April 2020) via a Twitter search Application Programming Interface (API) called from within R software the ‘twitterR package’[8]. We downloaded 10,000 Tweets per hashtag per day, consisting of a total of 210,000 randomly generated Tweets written in English. Each downloaded Tweet included information about the date the Tweet was created, status source (i.e. device used to post Tweet), screen name or user handle, the number of re-tweets (‘reposted tweets’), Tweet text and weblinks or URLs of the Tweet.

Tweets from all three hashtags were merged and pre-processed by removing duplicates including re-tweets, emoticons, emojis, punctuations, special characters and URLs. After pre-processing, a total of 39,726 Tweets were analysed. Trends in the volume of Tweets and frequencies across the 7-day period were described using line and bar graphs. Tweets were analysed for sentiments using the Bing sentiment lexicon from the R ‘tidytext’ package [9]. The Bing lexicon detects the sentiment of words through a dictionary lookup and then classifies words into binary “positive” or “negative” sentiments. All analyses including graphs were produced in R version 3.5.2.

Results

Tweet volume and frequency

After pre-processing, a total of 39,726 unique Tweets were identified across the three COVID-19 related hashtags. The volume of Tweets (Fig. 1) in each hashtag varied considerably across the 7-day period: #COVID19 (1,518 – 2,259), #SARSCoV2 (1,084 – 2,640), #Coronavirus (1,566 – 1,982). Trends in the volume of Tweets for #COVID19 and #SARSCoV2 were similar across the 7-day period, and both hashtags had a peak in the number of Tweets on the 3rd (Friday) and 6th (Monday) of April (Fig. 1).

Figure 1: Trends in the volumes of Tweets by hashtags

Tweets posted on these two days (See Points A, B, C and D in Fig 1) were further explored to gain insight into what may have triggered a heightened public response across the two hashtags. Table 1 provide some examples of Twitter discussions on both days.

Table 1: Examples of Tweet posted on the days with highest Tweet Volumes in #SARSCoV2 and #COVID-19

Date	#SARSCOV2	#COVID19
3rd of April	Point A	Point B
	Globally: 1,056,256 confirmed #coronavirus #covid19 cases; 55,770 fatalities = ~5.3% or 1 in ~19 deaths.	How long did it take to reach one million #COVID19 confirmed cases in the
	John Hopkins 4/3/2020 12:31AM PDT: #SARSCoV2 (virus) #COVID19 (disease): Total Confirmed 1,076,017 Total Deaths 58...	COVID-19 Updates (globally) Total cases: 1,093,103. Dead: 58,729 Recovered: 228,039 Last updated: April 03, 2020...
	Here's a study that everyone might want to read about miracle cure of chloroquine/hydroxychloroquine+azithromycin....	Another study shows the benefit of HCQ in #COVID19 Efficacy of hydroxychloroquine in patients with COVID-19:
	XXXXX on effectiveness of hydroxychloroquine: 'I think this is the beginning of the end of the pandemic...	Preach! THE FIX: XXXXX, an infectious disease specialist, on using hydroxychloroquine/azithromycin ...
6th of April	Point C	Point D
	We hope British Prime Minister Boris Johnson pulls through, hopefully without requiring intubation and ventilator....	UK PM Boris Johnson in intensive care with #COVID19 This virus can hurt anyone Im terrified for my sick, sufferin...
	So Boris has been moved to ICU. It's quite clear that #SARSCoV2 does not discriminate. Whilst I do not agree with a...	I was at a hospital where there were a few #coronavirus patients and I shook hands with everybody, you will be ple...

Results presented in Table 1 show that topics related to “Hydroxychloroquine” and “Hospitalisations of the British Prime minister” and “reaching 1 million cases of coronavirus globally” were topics similarly discussed by users of #COVID19 and #SARSCoV2 on both days.

In terms of frequency, Coronavirus related terms (e.g. “Coronavirus”, “covid19” and “sarscov2”) were the most frequently used terms across all three hashtags (Fig S1). Each had a frequency count of more than 6,000. The words “Pandemic”, “trump”, “people”, “Covid”, “coronavirus” were similarly tweeted across all

hashtags. Results of all three hashtags combined showed that “coronavirus”, was the most tweeted word, occurring more than 7000 times.

Figure 2: Sentiment of Tweets by hashtags

Sentiment analysis:

Results of the sentiment analysis (Fig 2.) showed that the majority of words used in the Tweets posted conveyed a negative sentiment towards the pandemic (#COVID19, 63.9%; #Coronavirus: 65.6%; #SARSCoV2 63.5%). The #Coronavirus had the largest proportion of words with a negative sentiment (65.6%). The words “trump” and “virus” were the most tweeted words that conveyed a negative and positive sentiment respectively (See Table S1).

Discussion

This exploratory study has shown that there is potential for using data generated from Twitter to understand general public opinion and sentiments towards the COVID-19 pandemic. We found that the volume of Tweets varied considerably across hashtags during the study period. The volume of Tweets across two hashtags (i.e. #COVID19 and #SARSCoV2) followed a similar pattern to each other, both experiencing a peak in Tweet volume on the 3rd and 6th of April 2020. A sharp increase in Tweet volume is an indicator of public interest or attention [5]. Further exploration of Tweets from the two hashtags on both days indicated that these similar spikes in Tweet volume may be linked to discussions on certain key events preceding those days. First, on the 2nd of April, 2020 it was announced that the confirmed cases of Coronavirus had surpassed 1 million globally based on data from Johns Hopkins University [10]. Second, on the 5th of April the UK Prime Minister was admitted to hospital for coronavirus symptoms. Another topic of interest during these dates, revealed by exploring tweets, was the use of ‘Hydroxychloroquine/ Azithromycin’ to treat COVID-19. These events may have generated greater interest and Twitter discussion, particularly amongst users of #COVID19 and #SARSCoV2 hashtags. These findings suggest that changes in Tweet volume or frequency could provide greater insight into how the public react to changing events during the coronavirus pandemic. Further research should examine how public sentiment varies with Tweet volumes.

We found that most words used within Tweets (i.e. over 60%) conveyed a negative sentiment or mood towards the pandemic. Our results also showed that #Coronavirus contained the most words with negative sentiment (65.6%). Our study, however, did not consider the context of Tweets or how words were used within Tweets. Therefore, it is difficult to conclude whether the large proportion of negative words identified within Tweets across hashtags actually reflected negative sentiments towards the pandemic during the 7-day period. Future studies should use more advanced text analytic techniques (e.g. Topic modelling, n-grams) to confirm these findings. In addition, using frequently identified ‘negative’ words as a basis for further exploration may provide further insight and inform a greater understanding of public sentiment about the pandemic.

Conclusion

The study has shown that there is scope for using Twitter data to efficiently and rapidly explore public sentiments and opinion about the current COVID-19 pandemic. Our results revealed some interesting insights from the analysis of Tweet trends and sentiments. However, greater consideration of the context of words or how words are used within Tweets is required.

Strength and limitations:

A major strength of this study is the use of publicly accessible Twitter data to rapidly explore general insights about public mood towards the COVID-19 pandemic.

Nevertheless, there are several limitations with this exploratory analysis. First, Tweets used in the analysis were randomly selected English Tweets derived from three popular hashtags and are limited to a 7-day period. Therefore, we must exercise caution in interpreting and generalising the findings from this study. Future work should consider using data from other coronavirus-related hashtags, including non-English Tweets and explore Tweets across an extended timeframe or at multiple time points. Second, we used a limited number of Tweets due to daily restrictions imposed by the Twitter search API. Future research should consider using a larger dataset to examine public sentiments and opinions with a search API that allows more flexibility in terms of the number of Tweets that can be downloaded. Third, we cannot verify the accuracy of the content analysed from Tweets because of the rise of misinformation or 'infodemic' as highlighted by the WHO [11]. Fourth, we did not have access to geolocated Tweets (i.e. geographical location of tweets). Thus, it is difficult to ascertain whether tweets are globally representative of public opinion. It would be helpful for future studies to capture the geographical location of tweets. Finally, we used limited categories for the analysis of sentiments (i.e. 'positive' or 'negative'). It is possible that words which convey a neutral sentiment may have been misclassified. Additionally, and importantly, within this analysis we did not consider the usage and context of the words within tweets. For example, the word 'positive' – may not necessarily indicate something good or positive (e.g. "tested positive" or "tests positive") but was translated to indicate positive sentiment due to the lexicon used (Table S1). Further research is needed to better understand the context around tweets by using lexicon that captures the entire content in a Tweet. In addition, supporting results of sentiment analysis with emoticons and emojis would add further contexts.

List Of Abbreviations

API - Application Programming Interface

COVID-19 - Coronavirus disease 2019

SARSCoV2 - Severe acute respiratory syndrome coronavirus 2.

- Hashtag

Declarations

Acknowledgements:

We acknowledge the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) for supporting this study.

Funding

This study is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College London. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Availability of data and materials:

The data that support the findings of this study are available on request from the corresponding author EC.

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Competing interests:

The authors have declared that no competing interests exist

Authors' contributions:

EC wrote the first draft of the manuscript, with significant input from WG and HJ. All authors critically reviewed the manuscript and approved the final version for submission.

References

1. Ortiz-Ospina, E., *The rise of social media* 2019.
2. Lee, J., et al., *Health Information Technology Trends in Social Media: Using Twitter Data*. Healthc Inform Res, 2019. **25**(2): p. 99-105.
3. Oscar, N., et al., *Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer's Disease Stigma on Twitter*. J Gerontol B Psychol Sci Soc Sci, 2017. **72**(5): p. 742-751.

4. Signorini, A., A.M. Segre, and P.M. Polgreen, *The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic*. PLoS One, 2011. **6**(5): p. e19467.
5. Chew, C. and G. Eysenbach, *Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak*. PLoS One, 2010. **5**(11): p. e14118.
6. Sinnenberg, L., et al., *Twitter as a Potential Data Source for Cardiovascular Disease Research*. JAMA Cardiology, 2016. **1**(9): p. 1032-1036.
7. Tavoschi, L., et al., *Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy*. Hum Vaccin Immunother, 2020: p. 1-8.
8. Gentry, J., *twitterR: R Based Twitter Client . R package version 1.1.9*. 2015.
9. Silge, J. and D. Robinson, *tidytext: Text Mining and Analysis Using Tidy Data Principles in R*. The Journal of Open Source Software, 2016. **1**.
10. CNBC. *Worldwide coronavirus cases reach 1 million, doubling in a week as death toll tops 50,000*. 2020 19/05/2020]; Available from: <https://www.cnbc.com/2020/04/02/worldwide-coronavirus-cases-reach-1-million-doubling-in-a-week.html>.
11. Zarocostas, J., *How to fight an infodemic*. The Lancet, 2020. **395**(10225): p. 676.

Supplementary Materials

Additional file 1. Figure S1. Frequency of Tweets by Hashtags.

Additional file 2: Table S1. Top 10 most frequently Tweeted words by Sentiments

Figures



Figure 1

Trends in the volumes of Tweets by hashtags

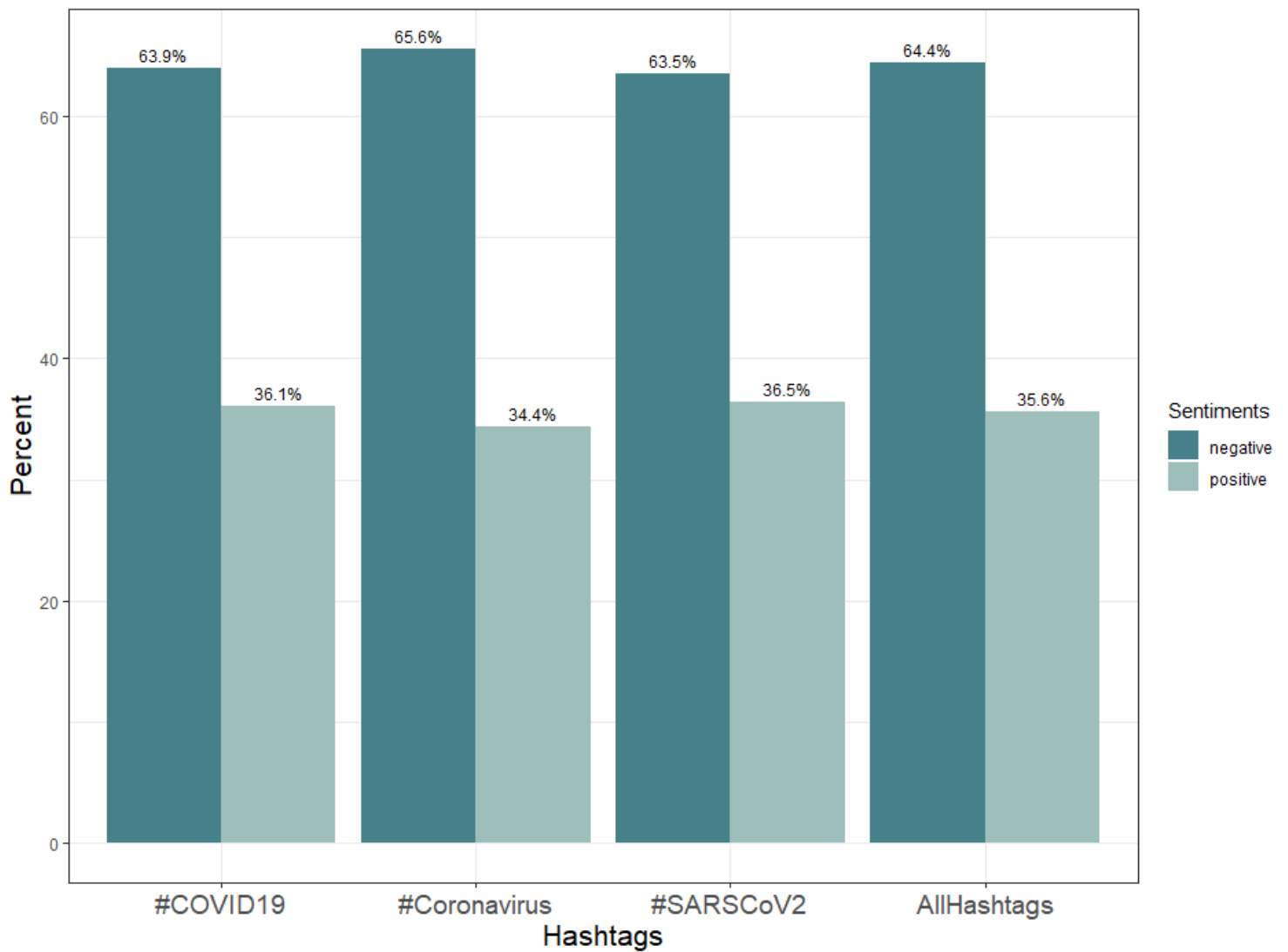


Figure 2

Sentiment of Tweets by hashtags

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.PNG](#)
- [TableS1.docx](#)