



A call for governments to pause Twitter censorship: using Twitter data as social-spatial sensors of COVID-19/SARS-CoV-2 research diffusion

Vanash M. Patel^{1,2} · Robin Haunschild³ · Lutz Bornmann⁴ · George Garas¹

Received: 14 July 2020 / Accepted: 15 December 2020 / Published online: 28 February 2021
© The Author(s) 2021

Abstract

In this study we determined whether Twitter data can be used as social-spatial sensors to show how research on COVID-19/SARS-CoV-2 diffuses through the population to reach the people that are affected by the disease. We performed a cross-sectional bibliometric analysis between 23rd March and 14th April 2020. Three sources of data were used: (1) deaths per number of population for COVID-19/SARS-CoV-2 retrieved from John Hopkins University and Worldometer, (2) publications related to COVID-19/SARS-CoV-2 retrieved from World Health Organisation COVID-19 database, and (3) tweets of these publications retrieved from Altmetric.com and Twitter. In the analysis, the number of publications used was 1761, and number of tweets used was 751,068. Mapping of worldwide data illustrated that high Twitter activity was related to high numbers of COVID-19/SARS-CoV-2 deaths, with tweets inversely weighted with number of publications. Regression models of worldwide data showed a positive correlation between the national deaths per number of population and tweets when holding number of publications constant (coefficient 0.0285, S.E. 0.0003, $p < 0.001$). Twitter can play a crucial role in the rapid research response during the COVID-19/SARS-CoV-2 pandemic, especially to spread research with prompt public scrutiny. Governments are urged to pause censorship of social media platforms to support the scientific community's fight against COVID-19/SARS-CoV-2.

Keywords Altmetrics · Twitter · Spatial maps · COVID-19 · SARS-CoV-2

✉ Vanash M. Patel
vanash.patel06@imperial.ac.uk

¹ Department of Surgery and Cancer, Imperial College London, 10th Floor, Queen Elizabeth the Queen Mother Wing, St. Mary's Hospital, London W2 1NY, UK

² Department of Colorectal Surgery, West Hertfordshire NHS Trust, Watford General Hospital, Vicarage Road, Watford, Hertfordshire WD18 0HB, UK

³ Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany

⁴ Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany

Introduction

Twitter is a social network created in 2006, that brings together hundreds of millions of users around its minimalist concept of microblogging, allowing users to post and interact with messages known as ‘tweets’ (https://about.twitter.com/en_us/company.html, 2020). Twitter has short delays in reflecting what its users perceive, and its principle of “following” users without obligatory reciprocity, together with a very open application programming interface, make it an ideal medium for the study of online behaviour (Grandjean 2016). Tweets can be used as ‘social sensors’, which is the concept of transforming a physical sensor in the real world through social media analysis. Tweets can be regarded as sensory information and Twitter users as sensors. Studies have demonstrated that tweets analysed as social sensors can provide insight into major social and physical events like earthquakes (Sakaki et al. 2010), sporting events (Takeichi et al. 2014), celebrity deaths (Sankaranarayanan et al. 2009), and presidential elections (Shamma et al. 2009). Twitter data contain location information which can be converted into geo-coordinates and spatially mapped. In this way, tweets can be used as social-spatial sensors to demonstrate how research diffuses within a population (Bornmann et al. 2020).

Researchers are increasingly using Twitter as a communication platform, and tweets often contain citations to scientific publications (Priem and Costello 2010). Twitter citations can form part of a rapid dialogue between users which may express and transmit academic impact and support traditional citation analysis. Twitter citations are defined as direct or indirect links from a tweet to a peer-reviewed scholarly article online (Priem and Costello 2010; Sakaki et al. 2010), and reflect a broader discussion crossing traditional disciplinary boundaries, as well as representing ‘attention, popularity or visibility’ rather than influence (Mas-Bleda and Thelwall 2016).

Coronavirus disease 2019 (COVID-19) is a novel infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The World Health Organization (WHO) declared the 2019–2020 coronavirus outbreak a Public Health Emergency of International Concern (PHEIC) [Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) 2020] on 30 January 2020 and a pandemic on 11 March 2020 (Ghebreyesus 2020).

We use Twitter data as social-spatial sensors to demonstrate how research on COVID-19/SARS-CoV-2 diffuses through the population and to investigate whether research reaches the people that are especially affected by the disease. Bornmann et al. (2020) demonstrated using HIV, tuberculosis and malaria as examples that Twitter data can be used for this purpose.

Summary of related studies

In our previous study, we have summarized previous research on Twitter and spatial analysis of online activities (Bornmann et al. 2020). As well as our previous study, which investigated how Twitter may be used as social-spatial sensors to track research diffusion in countries affected by HIV, tuberculosis and malaria, other studies have also shown how Twitter can be used to monitor public concerns with infectious diseases that cause epidemics and pandemics. Studies have used textual analysis of tweets from live Twitter chats to

determine public and institutional concerns about Ebola and Zika diseases (Glowacki et al. 2016; Lazard et al. 2015). More recently, textual analysis of tweets related to COVID-19/SARS-CoV-2 has shown that it is possible to understand the major public concerns including public fears and trending topics of the disease (Leelawat et al. 2020).

Methods

Dataset used

We used three sources of data in this study: (1) deaths per number of population for COVID-19/SARS-CoV-2, (2) publications related to COVID-19/SARS-CoV-2, and (3) tweets of these publications. All data was retrieved and analyzed between 23rd March and 14th April 2020. We used this time frame because COVID-19/SARS-CoV-2 had been declared a global pandemic by the WHO. By the end of March, all European countries and more than 150 countries worldwide had been affected by the disease.

Deaths

We used deaths per number of population as a measure of severity of the outbreak of the virus in countries and USA states. We used deaths per one hundred thousand population for country specific data, which was retrieved from Coronavirus Resource Center at John Hopkins University (Mortality Analyses 2020). We used deaths per one million population for US state specific data, which was retrieved from Worldometer, a provider of global COVID-19 statistics trusted by institutions such as the United Kingdom government and The Center for Systems Science and Engineering at Johns Hopkins University (Worldometer's COVID-19 data 2020).

Publications

We used the WHO COVID-19 database of global publications, which is the latest international multilingual scientific findings and knowledge on COVID-19 from searches of bibliographic databases, hand searching, and the addition of other expert-referred scientific articles (Global research on coronavirus disease (COVID-19) 2020). In total, 2413 publications were downloaded in CSV format from the WHO COVID-19 database of global publications on 23 March 2020, out of which 1941 publications had DOI information. Some data cleaning was performed for increasing the number of valid and removing invalid DOIs using R (R Core Team 2019). Strings like “dx.doi.org” and “doi:” before a DOI were removed. In the case of some publications, a DOI was located in the column “Accession Number”. In these cases, the DOI column was empty and the DOI from the column “Accession number” was used.

For some publications, multiple DOIs occurred in the same field delimited with whitespaces. A few manual checks revealed that sometimes one of the DOIs seemed to be related to research on coronavirus disease but others were not. We decided to remove all publications which contained whitespaces in the DOI field. After the cleaning procedure, some DOIs occurred multiple times. The R package *plyr* was used during the data cleaning process (Wickham 2011). Finally, we obtained a set of 1782 unique DOIs which occurred

only once in the dataset. The countries of the author's affiliations were downloaded via the Dimensions API (see <https://www.dimensions.ai/>). We found 1761 out of 1782 DOIs. The R packages *rjson* (Couture-Beil 2014) and *sqldf* (Grothendieck 2017) were used for parsing and aggregating the response from the Dimensions API.

Tweets

We used Altmetric.com application programming interface to extract tweet identifiers for any tweets which mentioned any of the publications on Twitter (Bonasio 2014). The tweet IDs of the tweets which mentioned any of the publications on Twitter were downloaded from the Altmetric.com API using R (R Core Team 2019) with the R packages *httr* (Wickham 2017a) and *RCurl* (Lang and the CRAN team 2018). The tweets were downloaded between 27 March and 07 April 2020 via the Twitter API using R (R Core Team 2019) and stored in a local SQLite database file using the R package *RSQLite* (Müller et al. 2017). Functions from the R package *DBI* were used for sending database queries (R Special Interest Group on Databases (R-SIG-DB) et al. 2018). The R package *ggplot2* was used for plotting the time evolution of tweets (Wickham 2016). The R package *tidyverse* (Wickham 2017b) was used for analysis of the Twitter user profiles. The R package *UpSetR* (Gehlenborg 2019) was used for plotting classifications of Twitter user profiles.

In total, 757,133 tweets were downloaded which mentioned 1561 DOIs (87.6% of the DOIs extracted from the WHO data set). In the case of 6065 tweets, no tweet text and no meta-data could be retrieved via the Twitter API. We analyzed the remaining 751,068 tweets. Not all Twitter users provide information about their geographical location (Sakaki et al. 2010; Wouters et al. 2019). Only 13 tweets contained precise geo-coordinates but 494,137 of those tweets contained some free-text user location information. We discarded the precise geo-coordinates and used only the user location information. One problem with the free-text user location information is that some users seem to become very imaginative. In order to reduce wrong location information, we needed to filter the location information for meaningful entries. In summary, we imported the city and country names from the Global Research Identifier Database (GRID, <https://grid.ac/>) for obtaining a whitelist of existing cities and countries and performed data cleaning of the free-text user location information (Bornmann et al. 2020).

The unique location strings were passed to the Google API via the R package *ggmap* if the location strings contained more than three characters (Kahle and Wickham 2013). The Google API returned among others precise geo-coordinates, country, and state names (if available) which were stored in a CSV file for plotting and statistical analysis. Overall, the geo-coordinates of 118,994 tweets could be determined and used for further analysis.

Statistical analysis

We used several Stata commands in this study (Crow and Gould 2013; Huebler 2012; StataCorp 2017). The most important Stata commands were *shp2dta* (Crow 2006) and *spmap* (Pisati 2007) to produce the Twitter maps. We additionally calculated Poisson regression models with number of tweets as dependent variable and COVID-19/SARS-CoV-2 deaths per number of population and number of publications as independent variables. We included another binary independent variable reflecting national censorship of Twitter in Iran and China (1 = national censorship). With count variables as dependent variables, Poisson regression models are indicated (Deschacht and Engels 2014; Hilbe 2014).

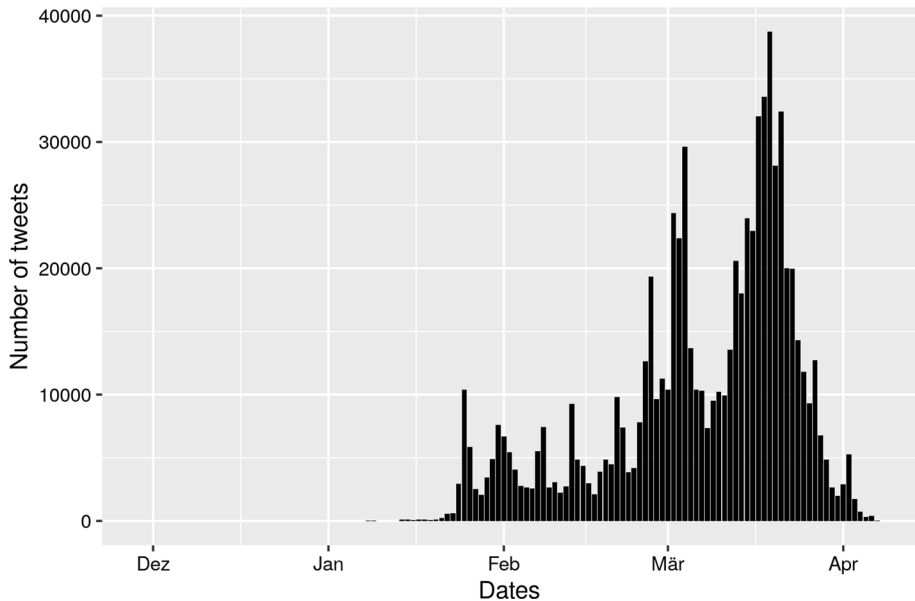


Fig. 1 Tweet timeline showing number of tweets per day since 17th November 2019

We focus on percentage changes in expected counts in the interpretation of the models (Long and Freese 2014). These percentages can be interpreted as follows: for a standard deviation increase in the death rate per population in a country (or US state), the increases in the expected tweet number in that country (or US state), holding the country’s (or US state’s) number of publications constant.

Results

The deaths per one hundred thousand population for countries ranged from 0 (Ethiopia) to 104 (San Marino). The deaths per one million population for USA states ranged from 2 (Wyoming) to 513 (New York). The total number of publications that were used in the analysis was 1761, and the total number of tweets that were used in the analysis was 751,068. Figure 1, which shows a barplot of the number of tweets per day since 17th November 2019 (date of the first known case of COVID-19/SARS-CoV-2), demonstrates exponential growth in twitter activity in March 2020.

The activity of Twitter sensors—which can be in the status “active” (i.e., tweeting) or not—on certain triggers (e.g., earthquakes or indications of influenza) can be measured. In this study, Twitter users’ function as social-spatial sensors by being aware of publications dealing with a certain disease. Since one can expect that the interest in publications on certain diseases increases, when the user is located in regions with many cases of illness, Twitter rates and disease rates might correlate. This relationship can only be assumed, however, if the general public is active on Twitter (besides researchers) and tweets about scholarly publications. Altmetric.com uses a popular classification system for Twitter users which is not ideal for our study because the analyses cannot be limited to only those Twitter users for whom we have geographical location information, and

Table 1 Twitter user classification scheme proposed by Toupin et al. (2019)

User	Profile
Faculty and students	Higher education or the realm of research
Communicators and journalists	Transmission of information at higher scale (e.g., media, arts, literature)
Professionals	Engaging with research publications relevant to their job (e.g., conservation manager)
Political	Engaging with research publications with political interest (e.g., through activism or as part of governmental jobs)
Personal	Self-describe themselves using personal interests (e.g., in cats or dogs)
Institutions and organizations	Represent a group of people
Bots	Use keywords related to automated activity
Journals and publishers	Represent journals or scientific publishers

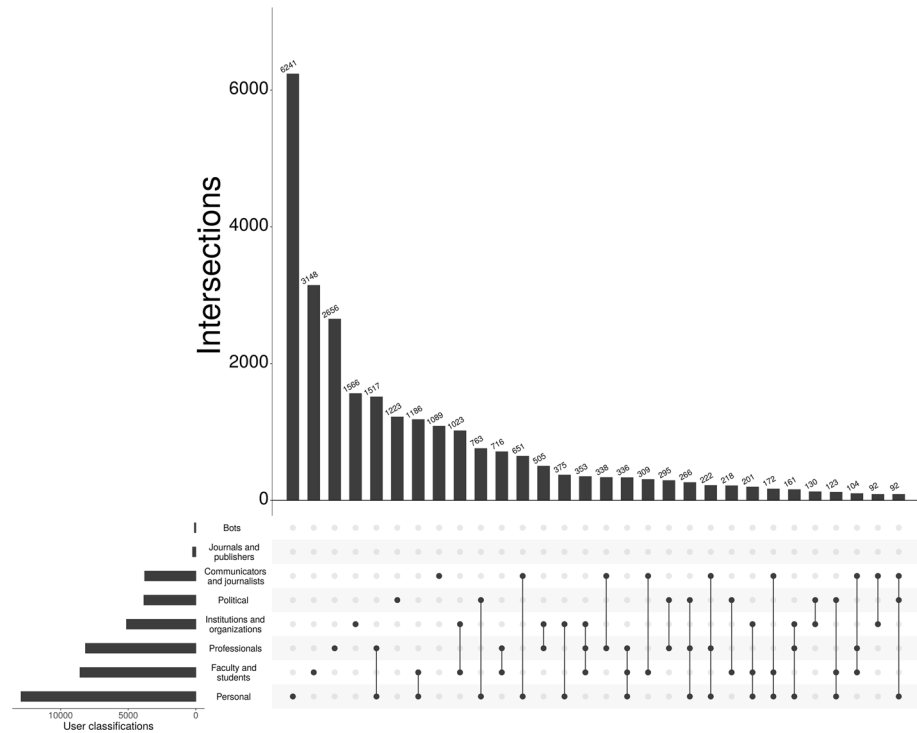


Fig. 2 Twitter user classification for tweets on publications dealing with COVID-19/SARS-CoV-2. The graph is restricted to the set of profiles with geographical location information that could be converted into geo-coordinates

Altmetric.com defines the group ‘members of the public’ as people who do not tweet links to scholarly publications. In order to receive information on the people tweeting on COVID-19/SARS-CoV-2 research, we used the classification scheme proposed by Toupin et al. (2019) (Table 1) and a modified version of the R code provided by Toupin (2020).

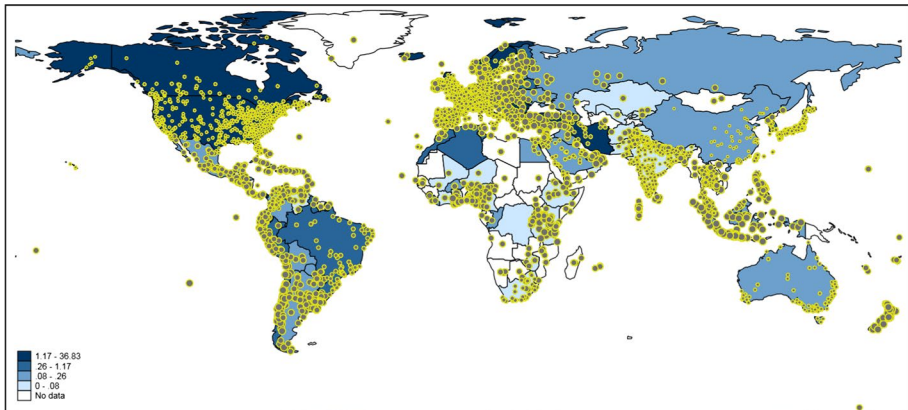


Fig. 3 Tweeting on publications dealing with COVID-19/SARS-CoV-2 worldwide. Each tweet is inversely weighted with the number of publications published by authors in the corresponding country: the larger the dots, the smaller the research activity. The countries are colored according to the national deaths per one hundred thousand population. For some countries, e.g. Greenland, no data are available. Countries such as China and Iran block internet access to Twitter or its content (Mortality Analyses 2020)

The result of the analysis is displayed in Fig. 2, which shows the number of tweets per user classification for tweets on publications dealing with COVID-19/SARS-CoV-2. The visualization reveals that many profiles (and their tweets) are assigned to more than one classification. There seems to be a large proportion of members of the general public in the dataset of this study. The largest proportion of Twitter user profiles belongs to the classification “Personal”. The second largest classification is “Faculty and students”. Other classifications that could be understood as members of the general public, too, (e.g., “Professionals”, “Institutions and organizations”, and “Political”) also contribute with sizable proportions. The classification with the lowest number of profiles is “Bots”.

Mapping worldwide data

Figure 3 shows worldwide Twitter activity referring to publications dealing with COVID-19/SARS-CoV-2. The underlying blue-colored scheme visualizes national deaths per number of population. The map is intended to show whether COVID-19/SARS-CoV-2 research reaches regions with many COVID-19/SARS-CoV-2 deaths: does the number of COVID-19/SARS-CoV-2 cases correlate with the number of tweets on COVID-19/SARS-CoV-2 publications?

One of the problems with Twitter data in the context of this study is that Twitter activity is generally high where more research is done (e.g., Western Europe or the Boston area in Fig. 1). Since this is not the activity which we intended to measure, we inversely weighted the size of each tweet on the map by the number of publications in that country [i.e., $1/\log(\text{number of publications})$]. Thus, the more publications’ authors are located in a country, the smaller the size of the tweet dot is (see here Ginsberg et al. 2009; Sinnenberg et al. 2017). We assume that large dots reflect tweets of people not doing research or not being a publisher/publishing organization (but might be personally confronted with COVID-19/SARS-CoV-2).

The map in Fig. 3 might show the expected result that high Twitter activity is related to high numbers of COVID-19/SARS-CoV-2 deaths. However, it is not completely clear

Table 2 Coefficients of a Poisson regression model with number of tweets as dependent variable (n = 111 countries)

Independent variable	Coefficient	Standard error	Percentage change in expected count
Deaths per number of population	0.0285***	0.0003	19.7
Number of publications	0.0152***	2.82×10^{-5}	155.5
National censorship of Twitter	-8.6662***	0.0376	
Constant	6.3950***	0.0040	

*** $p < 0.001$

whether this conclusion can be drawn, since there are several countries with high Twitter activity and high publication output (e.g., Western Europe and the Boston region). For some regions on the map, the extent of Twitter activity is difficult to interpret since tweet dots might overlap (especially those with larger sizes). To have a conclusive answer on the relation between Twitter activity and publication output, we additionally calculated Poisson regression models with number of tweets as dependent variable and deaths per number of population and number of publications as independent variables. In order to control the influence of research activity on Twitter activity, the number of publications has been considered as a second independent variable in the model.

The results are shown in Table 2. The coefficients of deaths per number of population and number of publications are statistically significant. The percentage changes in expected counts reveal that deaths per number of population and Twitter activities are related in fact: for a standard deviation increase in the national deaths per number of population, the expected number of tweets in that country increases by 19.7 percentage points, holding the country's number of publications constant. The results in Table 2 further show that the influence of the number of publications is significantly higher than that of deaths per number of population.

Mapping United States of America (USA) data

We did not only use the Twitter data as social-spatial sensors to investigate global trends, but also on a single country. We chose to use the USA as an example because it is one of the most populous nations, has the highest twitter activity and is the most prolific publisher of high-quality science (Garas et al. 2019).¹ Figure 4 shows publication-based Twitter activity dealing with COVID-19/SARS-CoV-2 in the USA. The blue-colored scheme presents the deaths in the USA states per one million population. The map might show that the deaths in the USA states are in fact related to the number of tweets on COVID-19/SARS-CoV-2 publications. However, there are several USA states with high Twitter activity and high publication output (e.g., the Boston region).

We calculated Poisson regression models with deaths per number of population and number of publications as independent variables and number of tweets as dependent variable. Table 2 reports the results. The results are based on 49 USA states (out of 51, Alaska and Hawaii were excluded) since only USA states with at least one tweet were considered. The

¹ <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries>.

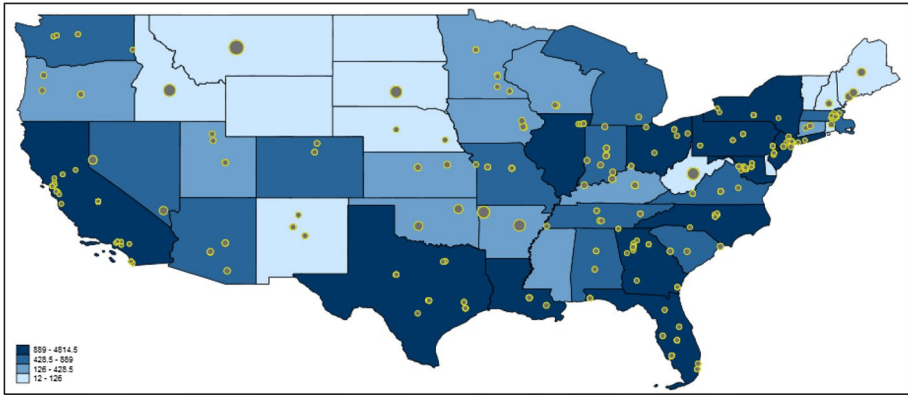


Fig. 4 Tweeting on publications dealing with COVID-19/SARS-CoV-2 in the USA. Each tweet is inversely weighted with the number of publications published by authors in the corresponding USA state: the larger the dots, the smaller the research activity. The USA states are colored according to their deaths per one million population (Worldometer’s COVID-19 data 2020)

Table 3 Coefficients of Poisson regression model with number of tweets as dependent variable (n=49 USA states)

Independent variable	Coefficient	Standard error	Percentage change in expected count
Deaths per number of population	-0.0013***	0.0001	-10.6
Number of publications	0.0744***	0.0005	105.5
Constant	5.1767***	0.0114	

*** $p < 0.001$

percentage changes in expected counts in Table 3 point out that deaths per number of population and Twitter activities are negatively correlated: for a standard deviation increase in the deaths per number of population of a USA state, the expected number of tweets in that state decreases by 10.6 percentage points, holding the USA state’s number of publications constant. The results in Table 3 further show that the influence of the number of publications is significantly greater than that of the deaths per number of population (and positive). In the USA states, there is a strong dependency of Twitter data on the number of publications.

The scatter plot of worldwide data in Fig. 5 demonstrates that at the time of the analysis the USA was an outlier because of lower national deaths per number of population and higher numbers of publications and tweets, when compared to other countries that were significantly impacted by COVID-19/SARS-CoV-2 (e.g., UK, France, Spain, and Italy).

Discussion

This study demonstrates that Twitter data can be used as social-spatial sensors to monitor research diffusion in a global pandemic using COVID-19/SARS-CoV-2 as an example. Our results suggest that novel research on COVID-19/SARS-CoV-2 publicised through Twitter reaches populations that are concerned about the disease.

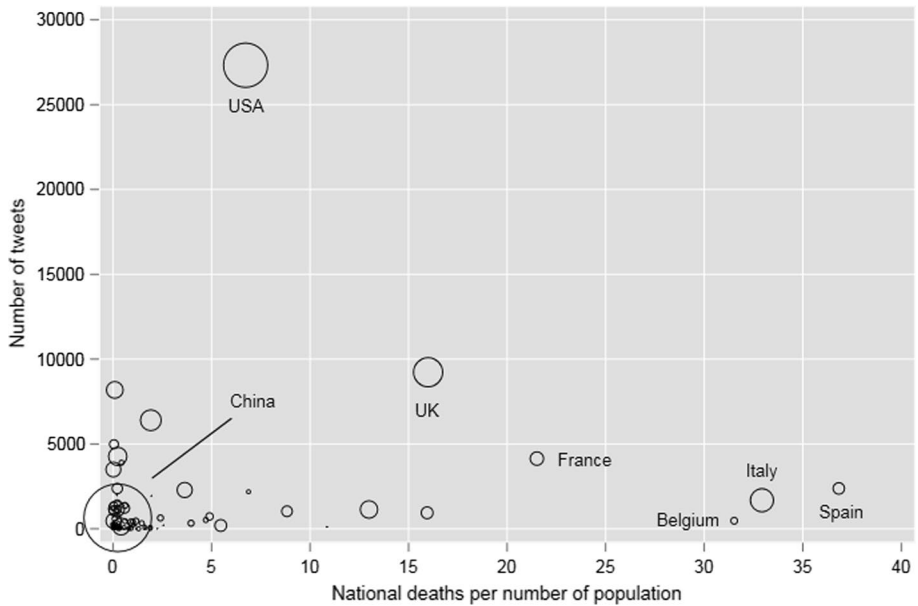


Fig. 5 Scatter plot of worldwide data showing deaths per number of population and number of tweets (markers for countries are weighted by the national numbers of publications)

Social media can be an effective tool for broadcasting research both within and beyond the academic community (Pulido et al. 2018). Twitter is one of the best social media platforms for sharing scientific research and knowledge because it allows users to post links of recent publications, write a short statement about the research topic and tag keywords with hashtags, so that people who are interested in the research are more likely to see the post (Mandavilli 2011). As well as promoting scientific research, Twitter and other social media platforms can scrutinise research in public, often within hours rather than years, unearthing poor quality inaccurate work (Mandavilli 2011). Governments and research institutions worldwide support a rapid research response to improve understanding of COVID-19/SARS-CoV-2, including the development and testing of therapies and vaccines (Statement from the leaders of the G7 nations 2020).

Our study shows that Twitter can play a vital role in the rapid research response, especially to disseminate research with swift peer review. Our study shows exponential use of Twitter as the intensity of the outbreak has increased. Over 80% of publications extracted from the WHO COVID-19 database have been cited on Twitter, which is nearly seven times higher than previous studies analysing Twitter data in biomedical sciences (Haustein et al. 2014). Each COVID-19 publication has been tweeted on average 425 times, which is significantly higher than our previous work analyzing Twitter activity of single infectious diseases (on average publications related to Human immunodeficiency virus were tweeted 7 times, tuberculosis were tweeted 8 times, and malaria tweeted 9 times) (Bornmann et al. 2020).

Countries such as China and Iran have blocked Twitter, as well as other social media platforms (Mohammadi et al. 2018). This is reflected in our mapping of worldwide tweet data related to research on COVID-19/SARS-CoV-2 (Fig. 2). The COVID-19/SARS-CoV-2 originated from Wuhan in China's Hubei province, which quickly became

the epicentre for China's outbreak, followed by a new epicentre in Iran. Both countries have seen a rapid rise in scientific output over the last two decades and their research (Ataie-Ashtiani 2016), coupled with thousands of reported cases, on COVID-19/SARS-CoV-2 has led them to a better understanding of the novel, fast moving virus. However, censorship of social media may have stifled research dissemination and more importantly avoids swift public scrutiny. This may adversely affect the global fight against the disease. Our study suggests that governments should consider relaxing censorship of social media at times of global crisis, such as the COVID-19/SARS-CoV-2 pandemic. Moreover, allowing the public greater access to platforms such as Twitter during a global pandemic can aid the scientific community's fight against misinformation and pseudoscience (Caulfield 2020).

The USA appears to be an outlier in the worldwide data and the country specific data shows that the USA has a different relationship between tweets and deaths, both of which may be due to the pandemic reaching the USA later than most other countries in the Northern Hemisphere. Another explanation is the difference in geographical clusters of COVID-19/SARS-CoV-2 and research productivity in the USA. On one hand, New York state became the global epicentre of the pandemic after the virus spread through Europe, and over a third of USA COVID-19/SARS-CoV-2 deaths have occurred in New York state, with the majority in New York City (Worldometer's COVID-19 data 2020). On the other hand, the USA is the most prolific publisher of high-quality science in the world, but the top-performing institutes are concentrated in Massachusetts, California and Maryland (Crew and Jia 2020).

Before concluding, it is important to consider the limitations of this study. This study focusses on Twitter although similar tools exist (e.g., Facebook). We do not expect other results with alternative tools, and the consideration of alternative tools in future studies might be an interesting addition. Although Twitter may be blocked by certain countries, people in these countries may use other microblogging platforms. For example, Sina Weibo is a Chinese microblogging site which is one of their most influential social network platforms, and we cannot determine whether research dissemination has occurred through Sina Weibo or similar platforms.

We have analysed tweets mentioning publications in a quantitative manner which does not account for the association of the tweet with the publication (i.e., a tweet may reference a valid study but claim it to be 'fake news' or have another negative overtone). We have not performed any thematic analysis of the tweets in terms of their content (e.g., are tweets referring to testing for COVID-19/SARS-CoV-2, therapies, or vaccines), or quality (e.g., are tweets referring to randomised controlled trials or letters). Since this study is based on publications from the international literature, people who do not know English cannot read the publications. Thus, many people with possible interest in COVID-19/SARS-CoV-2 publications cannot receive the scientific literature and tweets are not expectable. Which of the tweets on COVID-19/SARS-CoV-2 publications are 'only' retweets and not original tweets? Would a different handling of retweets yield different results? These are interesting questions which might be an interesting topic of further research.

Despite these limitations, our study has a number of strengths. We have used an evidence-based and robust methodology to clean and analyse data, as well as extracting data from several well-established databases containing real world evidence updated in real time (Bonasio 2014; Global research on coronavirus disease (COVID-19) 2020; Mortality Analyses 2020; Worldometer's COVID-19 data 2020) Our study comes at a very critical point in time, when a rapid research response is vital to develop therapies and vaccines to slow the COVID-19/SARS-CoV-2 pandemic and lessen the damage of

the disease. Our study utilising Twitter data as social-spatial sensors can serve as proof-of-concept for future studies on Twitter and the evolving pandemic.

Conclusion

COVID-19/SARS-CoV-2 began as a cluster of cases of pneumonia in Wuhan, Hubei Province, but the outbreak quickly progressed from an PHEIC to a pandemic, which highlights the dynamic process of the spread of an infectious disease (Ghebreyesus, 2020; Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) 2020) Our study has simply investigated a snapshot of the relationship between this pandemic, research outputs, and Twitter activity, but demonstrates the importance of how social media platforms can be crucial to spread research with rapid scrutiny, which may also impede the degree of misinformation. We urge governments to pause censorship of social media platforms such as Twitter during these unprecedented times to support the scientific community's battle against COVID-19/SARS-CoV-2.

Acknowledgements Meta-data for publications were downloaded via the Dimensions API. Twitter data were retrieved from the Altmetric.com API. Tweets with their location information were retrieved from the Twitter API. The authors thank Rodrigo Costas (CWTS), Vincent Lariviere (University of Quebec, Montreal), Remi Toupin (University of Quebec, Montreal), and Stacy Konkiel (Altmetric.com) for helpful discussions regarding the analysis of location information and profile classification of Twitter users.

Authors contributions VMP: Literature search, study design, data collection, data analysis, data interpretation. RH: Study design, data collection, data interpretation. LB: Study design, data collection, data analysis, data interpretation, figures. GG: Study design, data interpretation. All authors contributed to drafting the article or revising it critically for important intellectual content. All authors have approved the final version of the manuscript. All authors had full access to all of the data (including statistical reports and tables) in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding None.

Data availability The full data set and the statistical code can be obtained, upon request, from the corresponding author.

Compliance with ethical standards

Conflict of interest We have no conflict of interests.

Ethical approval Manuscript approved for submission by internal peer review at Max Planck Institute, Germany.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ataie-Ashtiani, B. (2016). Chinese and Iranian scientific publications: Fast growth and poor ethics. *Science and Engineering Ethics*, 23(1), 28–29. <https://doi.org/10.1007/s11948-016-9766-1>.
- Bonasio, A. (2014). A look at Mendeley readership statistics. Retrieved October 14, 2014, from <http://blog.mendeley.com/academic-features/a-look-at-mendeley-readership-statistics/>.
- Bornmann, L., Haunschild, R., & Patel, V. M. (2020). Are papers addressing certain diseases perceived where these diseases are prevalent? The proposal of Twitter data to be used as social-spatial sensors. <http://arxiv.org/abs/2004.13974>.
- Caulfield, T. (2020). Pseudoscience and COVID-19—we've had enough already. *Nature*. <https://doi.org/10.1038/d41586-020-01266-z>.
- Couture-Beil, A. (2014). rjson: JSON for R. Retrieved October 30, 2020, from <https://CRAN.R-project.org/package=rjson>.
- Crew, B. & Jia, H. (2020). Leading research institutions in the Nature Index 2020 Annual Tables. Retrieved October 30, 2020, from <https://www.natureindex.com/news-blog/leading-research-institutions-science-nature-index-annual-tables-twenty-twenty>.
- Crow, K. (2006). SHP2DTA: Stata module to converts shape boundary files to Stata datasets. Statistical Software Components S456718, Boston College Department of Economics, revised 17 Jul 2015.
- Crow, K. & Gould, W. (2013). Working with spmap and maps. Retrieved October 30, 2020, from <https://www.stata.com/support/faqs/graphics/spmap-and-maps/>.
- Deschacht, N., & Engels, T. E. (2014). Limited dependent variable models and probabilistic prediction in informetrics. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 193–214). Berlin: Springer.
- Garas, G., Cingolani, I., Patel, V. M., Panzarasa, P., Darzi, A., & Athanasiou, T. (2019). Evaluating the implications of Brexit for research collaboration and policy: A network analysis and simulation study. *British Medical Journal Open*, 9(9), e025025. <https://doi.org/10.1136/bmjopen-2018-025025>.
- Gehlenborg, N. (2019). UpSetR: A more scalable alternative to Venn and Euler diagrams for visualizing intersecting sets. R package version 1.4.0. Retrieved June 23, 2020, from <https://CRAN.R-project.org/package=UpSetR>.
- Ghebreyesus, T. A. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020: World Health Organization.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–U1014. <https://doi.org/10.1038/nature07634>.
- Global research on coronavirus disease (COVID-19). (2020). Retrieved March 23, 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>.
- Glowacki, E. M., Lazard, A. J., Wilcox, G. B., Mackert, M., & Bernhardt, J. M. (2016). Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. *American Journal of Infection Control*, 44(12), 1709–1711. <https://doi.org/10.1016/j.ajic.2016.05.025>.
- Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*, 3(1), 1171458. <https://doi.org/10.1080/23311983.2016.1171458>.
- Grothendieck, G. (2017). sqldf: Manipulate R data frames using SQL. Retrieved October 30, 2020, from <https://CRAN.R-project.org/package=sqldf>.
- Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4), 656–669. <https://doi.org/10.1002/asi.23101>.
- Hilbe, J. M. (2014). *Modelling count data*. New York, NY: Cambridge University Press. https://about.twitter.com/en_us/company.html. (2020).
- Huebler, F. (2012). Guide to creating maps with Stata. Retrieved February 10, 2020, from <https://huebler.blogspot.com/2012/08/stata-maps.html>.
- Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1), 144–161.
- Lang, D. T., & the CRAN team. (2018). RCurl: General network (HTTP/FTP/...) Client Interface for R. Retrieved March 27, 2020, <https://CRAN.R-project.org/package=RCurl>.
- Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's

- Ebola live Twitter chat. *American Journal of Infection Control*, 43(10), 1109–1111. <https://doi.org/10.1016/j.ajic.2015.05.025>.
- Leclawat, N., Tang, J., Saengtabtim, K., & Laosunthara, A. (2020). Trends of tweets on the coronavirus disease-2019 (COVID-19) pandemic. *Journal of Disaster Research*, 15(4), 530–533. <https://doi.org/10.20965/jdr.2020.p0530>.
- Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata* (3rd ed.). College Station, TX: Stata Press, Stata Corporation.
- Mandavilli, A. (2011). Peer review: Trial by Twitter. *Nature*, 469(7330), 286–287. <https://doi.org/10.1038/469286a>.
- Mas-Bleda, A., & Thelwall, M. (2016). Can alternative indicators overcome language biases in citation counts? A comparison of Spanish and UK research. *Scientometrics*, 109(3), 2007–2030. <https://doi.org/10.1007/s11192-016-2118-8>.
- Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. *PLoS ONE*, 13(5), e0197265. <https://doi.org/10.1371/journal.pone.0197265>.
- Mortality Analyses. (2020). Retrieved March 23, 2020, from <https://coronavirus.jhu.edu/data/mortality>.
- Müller, K., Wickham, H., James, D. A., & Falcon, S. (2017). RSQLite: ‘SQLite’ Interface for R. R package version 2.0. Retrieved June 22, 2020, from <https://CRAN.R-project.org/package=RSQLite>.
- Pisati, M. (2007). SPMAP: Stata module to visualize spatial data, Statistical Software Components S456812, Boston College Department of Economics, revised 18 Jan 2018.
- Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–4. <https://doi.org/10.1002/meet.14504701201>.
- Pulido, C. M., Redondo-Sama, G., Sorde-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PLoS ONE*, 13(8), e0203117. <https://doi.org/10.1371/journal.pone.0203117>.
- R Core Team. (2019). R: A language and environment for statistical computing (version 3.6.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved March 27, 2020, from <https://www.r-project.org/>
- R Special Interest Group on Databases (R-SIG-DB), Wickham, H., & Müller, K. (2018). DBI: R database interface.
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. Paper presented at the proceedings of the 19th international conference on world wide web.
- Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., & Sperling, J. (2009). TwitterStand: News in tweets. Paper presented at the Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009). Tweet the debates: understanding community annotation of uncollected sources. Paper presented at the proceedings of the first SIGMM workshop on Social media.
- Sinnenberg, L., Bottenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health*, 107(1), e1–e8. <https://doi.org/10.2105/AJPH.2016.303512>.
- StataCorp. (2017). *Stata statistical software: Release 15*. College Station, TX: Stata Corporation.
- Statement from the leaders of the G7 nations. (2020). Prime Minister’s Office, 10 Downing Street and The Rt Hon Boris Johnson MP.
- Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). (2020). World Health Organization.
- Takeichi, Y., Sasahara, K., Suzuki, R., & Arita, T. (2014). Twitter as social sensor: Dynamics and structure in major sporting events. Paper presented at the ALIFE14.
- Toupin, R. (2020). twitterprofiles. Retrieved June 22, 2020, from <https://github.com/toupinr/twitterprofiles>
- Toupin, R., Millerand, F., & Larivière, V. (2019). Scholarly communication or public communication of science? Assessing who engage with climate change research on Twitter. Paper presented at the 17th international conference on scientometrics and informetrics (ISSI 2019) with a special STI conference track, Rome, Italy.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1–29. <https://doi.org/10.18637/jss.v040.i01>.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- Wickham, H. (2017a). httr: Tools for working with URLs and HTTP. Retrieved June 22, 2020, from <https://CRAN.R-project.org/package=httr>.

- Wickham, H. (2017b). tidyverse: Easily install and load the ‘Tidyverse’. R package version 1.2.1. Retrieved June 22, 2020, from <https://CRAN.R-project.org/package=tidyverse>.
- Worldometer’s COVID-19 data. (2020). Retrieved March 23, 2020, from <https://www.worldometers.info/coronavirus/country/us/>.
- Wouters, P., Zahedi, Z., & Costas, R. (2019). Social media metrics for new research evaluation. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 687–713). Cham: Springer.