Journal of Travel Medicine

International Society of Travel Medicine
Established 1991
Promoting healthy travel worldwide

Journal of Travel Medicine, 2020, 1–3
doi: 10.1093/jtm/taaa120
Advance Access Publication Date: 23 July 2020
Rapid Communication

Rapid Communication

Geolocated Twitter social media data to describe the geographic spread of SARS-CoV-2

Donal Bisanzio DVM, PhD[1,2,*], Moritz U G Kraemer DPhil[3,4,5], Thomas Brewer MS[4], John S Brownstein PhD[3,4] and Richard Reithinger PhD[1]

[1]RTI International, Washington, DC, USA, [2]Epidemiology and Public Health Division, School of Medicine, University of Nottingham, Nottingham, UK, [3]Department of Paediatrics, Harvard Medical School, Boston, USA, [4]Computational Epidemiology Lab, Boston Children's Hospital, Boston, USA, and [5]Department of Zoology, University of Oxford, Oxford, UK

*To whom correspondence should be addressed. Tel: +44 775 426 61 55; Email: dbisanzio@rti.org

Submitted 10 May 2020; Revised 2 June 2020; Editorial Decision 3 June 2020; Accepted 16 July 2020

Key words: SARS-CoV2, COVID-19, Twitter, epidemiology, geospatial, pandemic, mobility

As of 1 August, 2020, 17,396,943 confirmed cases of coronavirus disease (COVID-19) have been reported since December 2019, including 675,060 deaths, in >225 countries.[1] On 11 March 2020, World Health Organization declared COVID-19 a pandemic. We show how geolocated Twitter data can be used to predict the spatiotemporal spread of reported COVID-19 cases at the global level from China to identify those areas at high risk of becoming secondary outbreaks.

We applied an analytical approach previously used to study dengue transmission dynamics[2]; an analysis during the early stages of COVID-19 predicted 74.1% of locations were cases would occur.[3] Briefly, we used a convenience sample of openly available, geotagged Twitter data from 2013 to 2015 to estimate 2019–2020 human mobility patterns in and outside of China; at a global scale, mobility has shown to be fairly stable over long period of time.[4] Human mobility patterns were estimated by analyzing the Twitter data from users who had: (i) tweeted at least twice on consecutive days from China, between 1 December 2013 and 15 February 2014 and 1 December 2014 and 15 February 2015; and (ii) left China following their second tweet during the time period under investigation. Users' movements were tracked for 30 consecutive days after leaving China. Publicly available COVID-19[5] case data as of 20 March 2020 were used to investigate the correlation among cases reported and locations visited by the Twitter user study cohort.

During the selected time window, the number of Twitter users tweeting from China was 9687, for a total 1 063 908 geolocated tweets (median = 54, interquartile range [IQR] = 26–120). Among these users, 4669 (48.1%) posted tweets outside of China (421 117 [39.6%] tweets; median = 33; IQR = 11–91), with 3215 users (68.8%) posting more than two tweets from China between 1 December and 15 February in either 2014 or 2015—our study cohort. During the 30 days after leaving China following their second tweet, these users posted tweets from 2381 cities in 140 countries, for a total of 13 739 visits (median = 6; IQR = 3–13 cities visited) Figure 1. The countries with the highest number of visiting cohort users were the USA (1494, 46.5%), Japan (484, 15.1%), UK (447, 13.9%), Germany (275, 8.5%), Turkey (242, 7.5%), Thailand (238, 7.4%), Russia (234, 7.3%), France (226, 7.1%), India (213, 6.6%) and Brazil (199, 6.2%). The most visited cities were London (109 users, 3.4%, UK), Singapore (105 users, 3.2%), Tokyo (104 users, 3.2%, Japan), Bangkok (85 users, 2.7%, Spain), Sydney (72, 2.3%, Australia), New York (66, 2.1%, USA), Los Angeles (62, 1.9%, USA), Dubai (59, 1.8%, United Arab Emirates), Moscow (52, 1.6%, Russia) and Paris (52, 1.6%, France). The Spearman's rank correlation coefficient (ρ) obtained when comparing the number of country-level Twitter user visits and reported COVID-19 cases by 20 March 2020 showed a high correlation ($\rho = 0.71$, $P < 0.01$) Figure 1.

Several locations we identified in our analyses, including London, Singapore, Tokyo and Bangkok, were also previously identified as possible locations for severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) spread in an analysis of using 2019 International Air Transport Association data.[6] We used geolocated tweets instead of other data such as flights, census surveys, internet traffic and mobile phone activity, as these do not necessarily allow to identify travellers' intermediate or final destinations (e.g. flight data only capture the flight route but not visited cities; mobile phone data do not capture overseas

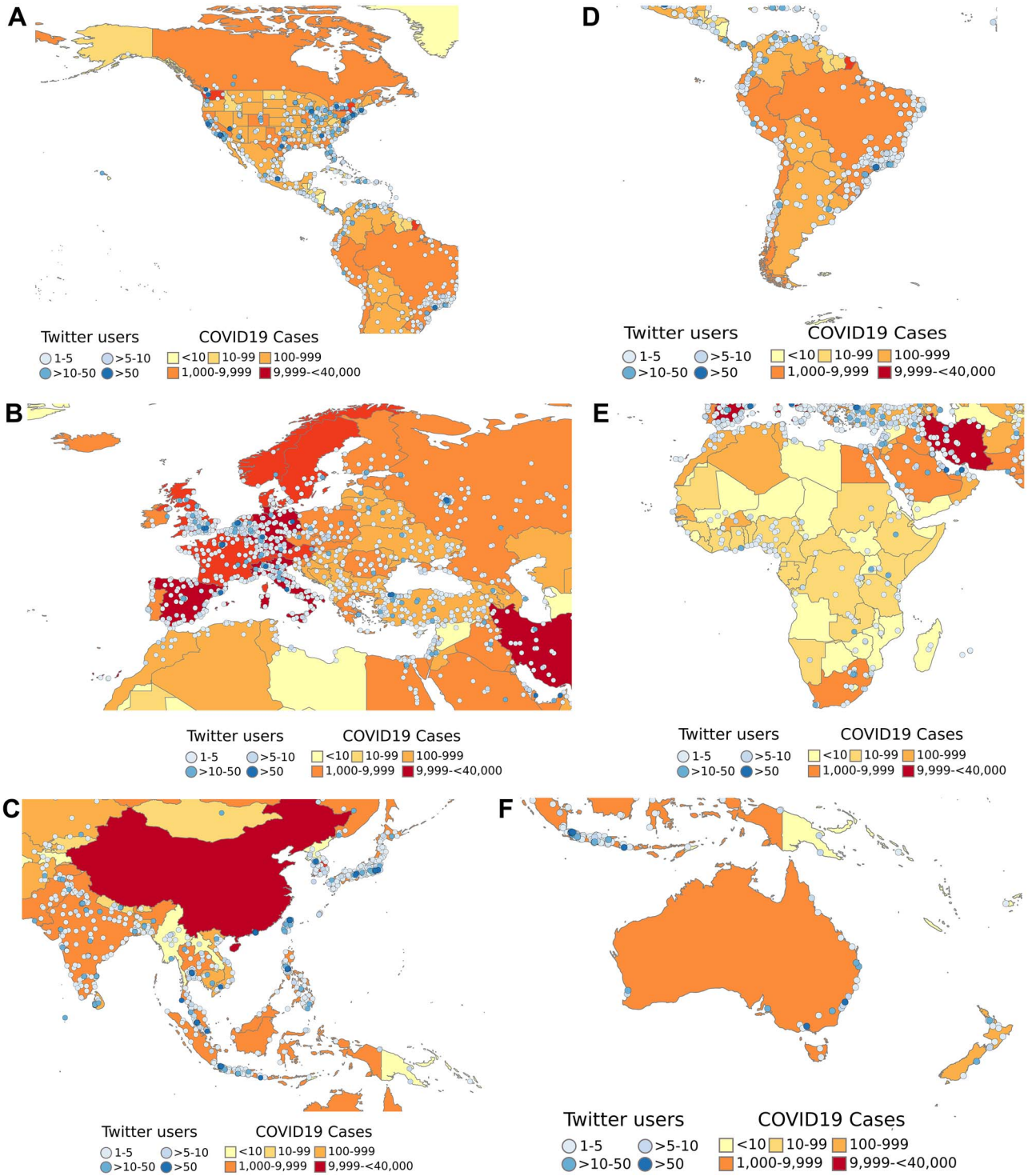

Figure 1. Location visited by the study cohort of Twitter users who were followed up for 30 days after having tweeted at least two times on consecutive days from Wuhan between 1 December 2013 and 15 February 2014 and 1 December 2014 and 15 February 2015. North and Central America (A), Europe (B), Asia (C), South America (D), Africa and Middle East (E) and Oceania (F).

trips).[6,7] Our analyses show that geolocated Twitter data can be used to describe the spread of a novel, highly transmissible agent such as SARS-CoV2 and identify areas at high risk of importation. This would allow public health authorities to develop appropriate response plans as well as start sensitizing public health providers and the population to the impending risk of exposure to such agent.

Conflicts of Interest

The authors declare no potential conflict of interest.

Funding

This study was conducted on institutional overhead funds. The opinions, results and conclusions reported in this paper are those

of the authors and are independent from funding sources of the authors' respective institutions and employers.

Author contributions

Study design, data analysis, data interpretation, writing (Donal Bisanzio; Richard Reithinger); data collection, data interpretation, writing (Moritz Kraemer, Thomas Brewer, John Brownstein).

References

1. WHO. *Coronavirus disease 2019 (COVID-19)*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200801-covid-19-sitrep-194.pdf?sfvrsn=401287f3_2 (1 August 2020, date last accessed)
2. Kraemer MUG, Bisanzio D, Reiner RC *et al*. Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan. *EPJ Data Science* 2018; **7**:16.
3. KraemerMUG D, Bogoch I *et al*.. Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global level. *Geospatial Health* 2020; **15**: 882.
4. Schneider CM, Belik V, Couronnné T, Smoreda Z, González MC. Unraveling daily human mobility motifs. *J R Soc Interface* 2013; **10**: 20130246.
5. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*w.
6. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Potential for global spread of a novel coronavirus from China. *J Travel Med* 2020; **27**. pii: taaa011. doi: 10.1093/jtm/taaa011
7. Lai S, Farnham A, Ruktanonchai NW, Tatem AJ. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. *J Travel Med* 2019; **26**:1–9.