BRIEF REPORT

Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of 2019 novel coronavirus at global level

Donal Bisanzio, DVM PhD;[1,2‡] Moritz U G Kraemer, PhD;[3,4,5] Isaac I Bogoch, MD;[6,7] Thomas Brewer;[4] John S Brownstein, PhD;[3,4] 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 Pediatrics, Harvard Medical School, Boston, USA.
[4] Computational Epidemiology Lab, Boston Children's Hospital, Boston, USA.
[5] Department of Zoology, University of Oxford, Oxford, UK.
[6] Department of Medicine, University of Toronto, Toronto, Canada.
[7] Divisions of General Internal Medicine and Infectious Diseases, University Health Network, Toronto, Canada.

[‡]Corresponding author:

Donal Bisanzio; RTI International, 701 13th Street NW, Suite 750, Washington, DC, United States 20005-3967; E-mail: dbisanzio@rti.org; Phone: +44 775 426 61 55

ABSTRACT

Importance: In December 2019 pneumonia cases of unknown etiological origin were reported in Wuhan, China. By January 30 2020, more than 8,235 confirmed cases of 2019 novel coronvirus (2019-nCoV) had been reported, including 170 deaths, with cases reported in Wuhan and other provinces in China, as well as in in 18 international locations outside of mainland China.

Objective: To predict the spatiotemporal spread of 2019-nCoV at global level.

Design: Human mobility patterns were estimated by using openly available geolocated Twitter social media data posted from November 1, 2013, to February 28, 2014, and from November 1, 2014, to February 28, 2015. Case data were used to investigate the correlation among cases reported during the current 2019-nCoV outbreak, locations visited by a study cohort of Twitter users, and airports with scheduled flights from Wuhan in China. Infectious disease vulnerability index (IDVI) data were obtained to identify the capacity of countries receiving travellers from Wuhan to respond to infectious disease threats.

Setting: Global.

Participants: Publicly available 2019-nCoV case data reported since December 2019.

Main Outcome(s) and Measure(s): Human mobility patterns as per Twitter users who had (1) tweeted at least twice on consecutive days from Wuhan between November 1, 2013, and January 28, 2014, and November 1, 2014, and January 28, 2015; and (2) left Wuhan following their second tweet during the time period under investigation.

Results: Our study cohort comprised 161 users posting more than two tweets within two consecutive days from Wuhan during the study period. Of these users, 133 (82.6%) posted tweets from 157 Chinese cities (1,344 tweets) during the 30 days after leaving Wuhan following their second tweet, with a median of 2 (IQR= 1–3) locations visited and a mean distance of 601 km (IQR= 295.2–834.7 km) traveled. Of our user cohort, 60 (37.2%) traveled abroad to 119 locations in 28 countries. Of the 82 2019-nCoV cases reported outside China as of January 30, 2020, 54 cases had known geolocation coordinates and 74.1% (40 cases) were reported less than 15 km (median = 7.4 km, IQR= 2.9–285.5 km) from a location visited by at least one of our study cohort's users. Countries visited by the cohort's users and which have cases reported by January 30, 2020, had a median IDVI equal to 0.74.

Conclusions and Relevance: Based on our analyses, we anticipate cases to be reported soon in the United Kingdom, Saudi Arabia and Indonesia, all countries to which more than one user from our study cohort travelled within 30 days after having tweeted a second time from Wuhan during our study period; additionally, countries with a moderate to low IDVI (i.e. ≤0.7) such as Barbados, Indonesia, Pakistan, and Turkey should be on high alert and develop 2019-nCOV response plans as soon as permitting.

Introduction

On December 30, 2019, pneumonia cases of unknown etiological origin were reported in Wuhan, China.[1] We now know that these cases were due to a coronavirus (2019 novel coronavirus [2019-nCoV]).

Coronaviruses are RNA viruses distributed broadly among humans, other mammals, and birds. Six coronavirus species are known to cause human disease.[2] Although most coronavirus infections are considered mild, two coronaviruses—severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV)—resulted in 10,590 cumulative cases in the past two decades, with mortality rates of 9.6% and 34.4%, respectively.[3,4] As with SARS-CoV and MERS-CoV, 2019-nCoV is probably of zoonotic origin and human-to-human transmission has been confirmed.[5] Early studies of hospitalized patients with confirmed 2019-nCoV reported that severe illness was seen in 32% of cases and case fatality rates ranged between 11–15%.[6] As of January 30, 2020, there have been 8,235 (8,124 [98.7%] in China) confirmed cases of 2019-nCoV, including 171 deaths. Cases have been reported in Wuhan and 31 other provinces in China, as well as in 18 countries, most recently in the Philippines, Sri Lanka, France, Germany, Finland, Canada, and the USA.[7,8]

Given the potential of SARS-CoV, MERS-CoV, and other viruses to rapidly spread nationally and globally by commercial air travel[9] we sought to characterize the possible spatiotemporal spread of 2019-nCoV by applying human mobility models and estimates derived from user activity of the social media platform Twitter.

Methods

Epidemiological data. We used publicly available 2019 n-CoV case data and aggregated to the town level (population > 50,000 people) on a weekly basis from December 31, 2019, to January 30, 2020.[8] In total, 8,235 confirmed cases recorded at the global level were included in our analyses. Data on a country's infectious disease vulnerability index (IDVI) was obtained from Moore et al. 2017;[10] the IDVI is a validated metric of a country's capacity to prepare for and respond to infectious disease threats.

Human mobility data and analytical approach. We applied an analytical approach previously used to study urban transmission dynamics of dengue.[11] Briefly, we used a convenience sample of openly available Twitter data from 2013–2015 to estimate human mobility patterns in 2019–2020 in Wuhan; at a global scale mobility has shown to be fairly stable over long periods of time.[12] Our database consists of global tweets (spatial search windows: latitude –90 to 90 latitude and -180 to 180 longitude) posted from November 1, 2013, to February 28, 2014, and from November 1, 2014, to February 28, 2015. The time period was chosen as it represents the months that the current 2019-nCoV outbreak occurred over until travel outside of Wuhan became severely restricted due to the quarantine imposed by the Chinese authorities. Each tweet has a unique user ID, latitude, longitude, and date (year, month, hour, second). Obtained Twitter data is restricted to 1% of tweets posted globally during that time period; [11] as previously shown, the amount of Twitter users with geo-located information would have represented 1% of the total global population in the study period.[12] Our analytical approach is illustrated in Figure 1.

Results

During the selected study period (November 1, 2013 to February 28, 2014, and November 1, 2014 to February 28, 2015), the number of Twitter users who posted tweets from Wuhan was 1,344 for a total 313,286 geolocated tweets (median = 6, interquartile range [IQR] = 1–30). Among the selected users, 307 (22.8%) posted tweets in locations outside Wuhan (24,649 [7.9%] tweets; median = 10; IQR= 3–38). with 161 users (12.0%) posting more than two tweets from Wuhan between November 1 and January 28—our study cohort. Of these users, 133 (82.6%) posted tweets from 157 Chinese cities (1,344 [71.9%] tweets) during the 30 days after leaving Wuhan following their second tweet (Figure 2A, Table 1), with a median of 2 (IQR= 1–3) locations visited and a mean distance of 601 km (IQR= 295.2–834.7 km) traveled. The most visited cities were Beijing (29 users, 18%), Shanghai (29 users, 18%), Guangzhou (25 users, 15.5%), and Nanjing (11 users, 6.8%).

As per Twitter activity of our user study cohort, 60 (37.2%) traveled abroad to a total 119 locations in 28 countries (Figure 2B, Table 1). The countries with the highest number of visiting users were the US (10, 16.3%), Thailand (7, 11.4%), Saudi Arabia (7, 11.4%), and Australia (6, 9.8%) (Table 1). The most visited cities were Bangkok (7 users), Mecca (5 users), London (5 users), Sydney (4 users), Kuala Lumpur (4 users), and Los Angeles (4 users); 15 users (25%) visited more than one city, with two users reaching a maximum of 5 cities visited.

For those 2019-nCoV cases reported by January 30, 2020, for which the city was available, we compared the distance to the locations visited by our study cohort and the airports connected to Wuhan. Locations visited by our cohort users were statistically closer to reported cases than airports with the median distance being 20.1 km (IQR= 3.6–95.4 km) and 75.9 km (IQR=25.1–187.8 km), respectively (Wilcoxon's rank test, $p<0.01$). Of the 82 cases reported outside China, 54 cases had known coordinates and 74.1% (40 cases) were reported less than 15 km (median = 7.4 km, IQR= 2.9–285.5 km) from a location visited by at least one of our cohort's users.

The countries visited by the cohort's users and which have cases reported by January 30, 2020, have an median IDVI equal to 0.74 (IQR = 0.67–0.89) (Table 1). In total, 14 countries (50%) outside China visited by the cohort's users have reported cases. Among the 10 countries visited by more than one users, 7 reported multiple cases before January 26, 2020 (Table 1).

Discussion & Conclusion

Using an analytical approach that has previously been used to understand local spread dynamics of dengue, we sought to characterize the spatiotemporal spread of 2019-nCoV. We decided to use geolocated tweets instead of data already used to predict 2019-nCoV spread such as flights, census surveys, internet traffic, and mobile phone activity,[14] as these approaches do not necessarily allow to identify travelers' intermediate or final destinations (e.g. flight data only capture the flight route but not visited cities; mobile phone data do not capture overseas trips).

Based on 2013–2014 and 2014–2015 Twitter user data, and given that major travel routes only marginally changed during the last 5 years, we analyzed the mobility of a cohort of people who had (1) tweeted at least twice from Wuhan between November 1 and January 28; and (2) left Wuhan between November 1 and January 28 following their second tweet. Our findings show that human mobility of these Twitter users is substantial, with a defined study cohort of 161 users travelling outside of Wuhan. Of these, 133 travelled to 157 locations in China and 60 travelled to 119 locations in 28 countries. Of the 157 locations within China, 87 (55.4%) had—as of January 30, 2019—reported confirmed cases; of the 5,930 2019-nCoV cases with known location reported within China, 4,176 (70.4%) occurred in a location visited by at least one of our cohort's users. Of the 119 overseas locations, 15 (12.6%) had—as of January 30, 2019—reported confirmed cases; similarly, of the 54 2019-nCoV cases reported outside China with known location, 40 (74.1%) occurred in locations visited by at least one of our cohort's users.

A limitation of our study is that using Twitter to model human mobility could be biased towards a population that has access to a smartphone and use of the application; while this may be true, we note that the same population is also likely to have greater economic means for mid -and long-distance travel, a critical factor if assessing the global spread of 2019-nCoV. It is also likely that access to smartphones and Twitter since 2015 by the population in Wuhan may have changed, but it is less clear whether the human mobility patterns would have changed significantly—an issue which needs further investigation.

On January 30, 2020, WHO declared 2019-nCov to be a public health emergency of international concern.[15] The current response to contain the 2019-nCoV outbreak is evolving daily: in China several major cities are quarantined, with severe limitations on people's movements; internationally, several airlines have cancelled flights to China and some countries (e.g., US, United Kingdom, Italy) are evacuating their nationals as well as screening travelers coming from China at major ports of entry.

Based on our analyses, we anticipate that several locations that have yet to report 2019-nCoV cases are expected to have cases or report cases soon (Table 1). Of immediate concern for outbreak containment are—besides all identified cities in China—locations in countries in Central and South East Asia, i.e. cities that are easily accessible via direct flights, by road or sea from Wuhan and other Chinese cities (Table 1). Globally, we anticipate cases to be reported soon in the United Kingdom, Saudi Arabia and Indonesia, all countries where more than one user from our study cohort travelled to within 30 days after having tweeted a second time from Wuhan during our study period; additionally, countries with a moderate to low IDVI (i.e. ≤0.7) such as Barbados, Indonesia, Pakistan, and Turkey should be on high alert and develop 2019-nCOV response plans as soon as permitting.

References

1. WHO. Novel coronavirus – China. Jan 12, 2020. http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ (accessed January 31, 2020).
2. Su S, Wong G, Shi W, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol 2016; 24:490-502.
3. WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. Dec 31, 2003. https://www.who.int/csr/sars/country/table2004_04_21/en/ (accessed January 31, 2020).
4. WHO. Middle East respiratory syndrome coronavirus (MERS-CoV). November, 2019. http://www.who.int/emergencies/mers-cov/en/ (accessed January 31, 2020).
5. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020 Jan 24. pii: S0140-6736(20)30154-9.
6. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020; https://doi.org/10.1016/S0140-6736(20)30183-5
7. WHO. Novel Coronavirus(2019-nCoV) Situation Report – 10. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480_2 (accessed January 31, 2020).
8. http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-available-data/337/3
9. Findlater A, Bogoch, II. Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel. Trends Parasitol. 2018; 34: 772-83.
10. Moore M, Gelfeld B, Okunogbe A, Paul C. Identifying future disease hot spots: infectious disease vulnerability index." Rand health quarterly 2017; 6: .
11. 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.
12. Schneider CM, Belik V, Couronnné, E Smoreda Z, González MC. Unraveling Daily Human

Mobility Motifs. J R Soc Interface. 2013; 10: 20130246.

13. Lenormand M, Picornell M, Cantú-Ros OG, et al. Cross-checking different sources of mobility information. PloS One 2014; 9:

14. Lai S, Bogoch II, Watts A, Khan K, Li Z, Tatem A. Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China. https://www.worldpop.org/resources/docs/china/WorldPop-coronavirus-spread-risk-analysis-v1-25Jan.pdf (accessed January 31, 2020)

15. https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed January 31, 2020)

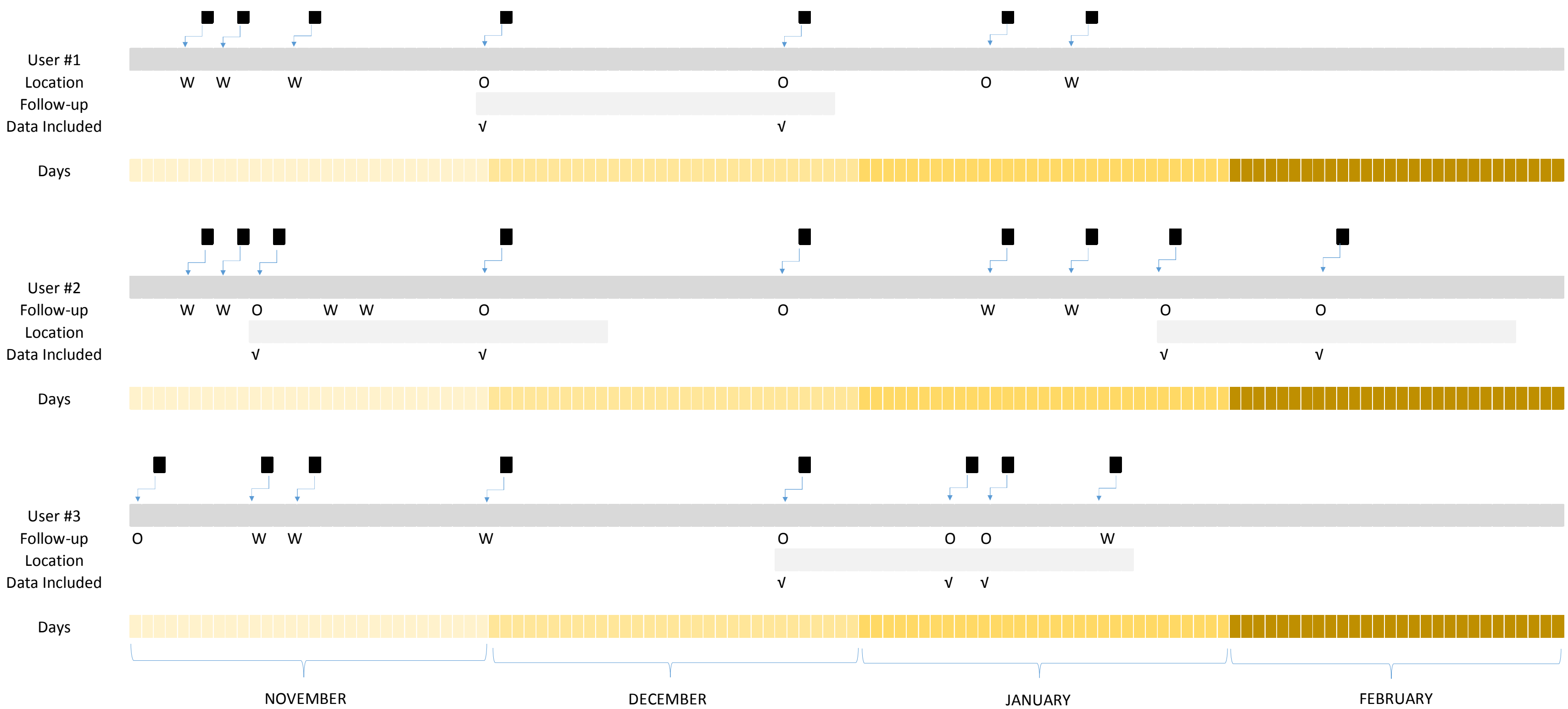
Table 1. Locations visited by the study cohort of Twitter users who were followed-up for 30 days after having tweeted at least two times on consecutive days from Wuhan between November 1, 2013, and February 28, 2014, and November 1, 2014, and February 28, 2015. The table reports: (1) the visited countries; (2) the number of cohort users traveling within the identified country; (3) the number of major cities (population > 50,000 people) visited by cohort users in each identified country; (4) the country IDVI; and (5) the date of first 2019-nCoV case reported.

Country	Number of users	Visited cities		IDVI	Date of first 2019 n-CoV case reported[a]
China	135	157	[Not listed]	0.663	December 30, 2019
USA	10	16	Allen, Atlanta, Chicago, Houston, Grand Prairie, Los Angeles, Mesquite, New York, Palo Alto, Pasadena, Richardson, San Diego, Santa Monica, San Mateo, Toledo, Washington DC	0.924	January 16, 2020
Saudi Arabia	7	4	Al-Madinah, Jiddah, Mecca, Riyadh	0.736	/
Thailand	7	8	Ayutthaya, Bangkok, Khlong Luang, Lam Luk Ka, Pak Kret, Phra Pradaeng, Samut Prakan, Saraburi	0.713	January 5, 2020
Australia	6	5	Brisbane, Geelong, Gold Coast, Melbourne, Sydney	0.912	January 15, 2020
Japan	5	20	Akita, Aomori, Beppu, Chitose, Dazaifu, Hachioji, Hakodate, Hino, Iwamizawa, Kitahiroshima, Musashino, Nagaoka, Oita, Saga, Sagamihara, Sakata, Sapporo, Tokyo, Tomakomai, Tomisato	0.926	January 3, 2020
UK	5	9	Cheadle, Doncaster, Edinburgh, Esher-Molesey, London, Manchester, Sheffield, Staines, Woking-Byfleet	0.89	/
Malaysia	4	9	Banting, George Town, Kajang-Sungai Chua, Klang, Kuala Lumpur, Petaling Jaya, Seremban, Subang Jaya, Sungai Ara	0.761	January 25, 2020*
Canada	3	5	Edmonton, Hamilton, Saint Catharines-Niagara, Toronto, Vancouver	0.973	January 22, 2020

Indonesia	3	8	Bandung, Ciamis, Cibeureum, Kadungora, Klaten, Sukabumi, Tangerang, Tasikmalaya	0.562	/
Singapore	3	1	Singapore	0.877	January 21, 2020
Barbados	1	1	Bridgetown	0.681	/
Brazil	1	7	Cacapava, Caieiras, Cotia, Diadema, Franco da Rocha, Guarulhos, Sao Paulo	0.716	/
Cambodia	1	1	Siem Reab	0.355	January 26, 2020
France	1	1	Paris	0.855	January 18, 2020
India	1	1	Bommanahalli	0.499	January 30, 2020*
Ireland	1	2	Dublin, Limerick	0.906	/
Italy	1	2	Modena, Verona	0.821	/
Mexico	1	1	Mexicali	0.734	/
New Zealand	1	3	Auckland, Christchurch, Wellington	0.916	/
Pakistan	1	2	Faisalabad, Lahore	0.308	/
Philippines	1	1	Davao	0.544	January 30, 2020*
Puerto Rico	1	2	Carolina, San Juan	0.924	/
Spain	1	3	Barakaldo, Bilbao, Getxo	0.875	/
Taiwan	1	1	Taichung	0.709	/
Turkey	1	4	Bozuyuk, Eskisehir, Istanbul, Sultanbeyli	0.677	/
United Arab Emirates	1	1	Dubai	0.765	January 29, 2020*
Vietnam	1	1	Ho Chi Min City	0.626	January 17, 2020

'/' no reported cases; [a] the date of onset symptoms; *confirmation date; IDVI, infectious disease vulnerability index

Figure 1. Analytical approach with Twitter activity of three illustrative users. Obtained Twitter database was filtered to only include users who posted at least two tweets on consecutive days within the city of Wuhan between November 1, 2013, and February 28, 2014, and November 1, 2014, and February 28, 2015, to ensure that the user was physically in Wuhan for at least 24 hrs. To characterize the possible spatiotemporal spread of 2019-nCoV, we then followed-up the Twitter activity of these users for 30 days post second tweet and determined whether these users travelled outside of Wuhan; we chose this follow-up period as we presumed that it would cover any 2019-nCoV pre-patent period if exposure would have happened prior to the users' second tweet. Using the geographic fingerprint of users' tweets, we estimated the locations visited by each user included in the study cohort by linking all tweets to the closest city. For movement of users within China, we also calculated the mean distance from Wuhan by averaging the maximum distance of each user based on their Twitter activity and the geographic fingerprint of their tweets. We used the Wilcoxon's rank test to compare the distance of visited locations and major airports connected to Wuhan from confirmed 2019-nCoV cases with known location (significance threshold set to $p<0.05$).



■ : user Twitter activity. Location: W, Wuhan; O, outside of Wuhan. Follow-up: light grey, 30 day follow-up period. Data included in the analyses: √, yes.

Figure 2A. South East Asia locations 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 November 1, 2013, and February 28, 2014, and November 1, 2014, and February 28, 2015. The figure includes airports with scheduled flights from Wuhan, locations of reported cases, and IDVI of countries visited by the study cohort.

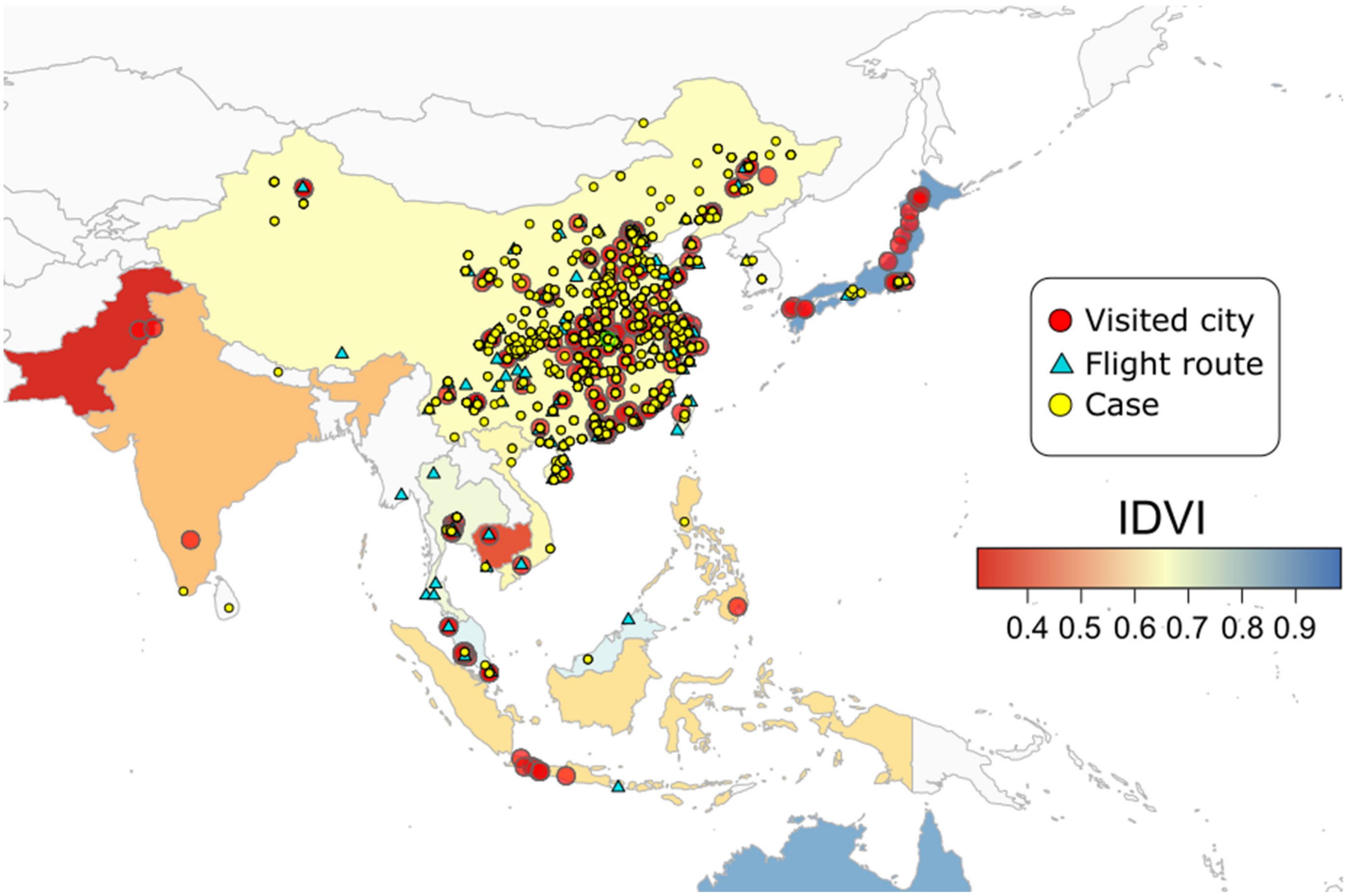


Figure 2B. Location visited by visited by the study cohort of Twitter users who were followed-up for 30 days after having tweeted at least two times on consecutive days from Wuhan between November 1, 2013, and February 28, 2014 and November 1, 2014, and February 28, 2015. The figure includes airports with scheduled flights from Wuhan, locations reporting cases, and IDVI of countries visited by the study cohort.