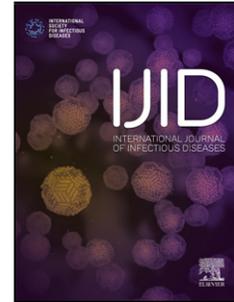


# Journal Pre-proof

First, second and potential third generation spreads of the COVID-19 epidemic in mainland China: an early exploratory study incorporating location-based service data of mobile devices

Bisong Hu, Jingyu Qiu, Haiying Chen, Vincent Tao, Jinfeng Wang, Hui Lin



PII: S1201-9712(20)30355-6

DOI: <https://doi.org/10.1016/j.ijid.2020.05.048>

Reference: IJID 4226

To appear in: *International Journal of Infectious Diseases*

Received Date: 14 March 2020

Revised Date: 8 May 2020

Accepted Date: 13 May 2020

Please cite this article as: Hu B, Qiu J, Chen H, Tao V, Wang J, Lin H, First, second and potential third generation spreads of the COVID-19 epidemic in mainland China: an early exploratory study incorporating location-based service data of mobile devices, *International Journal of Infectious Diseases* (2020), doi: <https://doi.org/10.1016/j.ijid.2020.05.048>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

## First, second and potential third generation spreads of the COVID-19 epidemic in mainland China: an early exploratory study incorporating location-based service data of mobile devices

Bisong Hu<sup>a,b</sup>, Jingyu Qiu<sup>c</sup>, Haiying Chen<sup>d</sup>, Vincent Tao<sup>c,\*</sup>, Jinfeng Wang<sup>b,\*</sup>, Hui Lin<sup>a,\*</sup>

<sup>a</sup>School of Geography and Environment, Jiangxi Normal University, Nanchang, China

<sup>b</sup>State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

<sup>c</sup>Wayz AI Technology Company Limited, Shanghai, China

<sup>d</sup>Nanchang Center for Disease Control and Prevention, Nanchang, China

\*Corresponding authors.

E-mail addresses:

hubisong@jxnu.edu.cn (B. Hu), jingyu.qiu@wayz.ai (J. Qiu), nccdchy@126.com (H. Chen), vincent.tao@wayz.ai (V. Tao), wangjf@reis.ac.cn (J. Wang), huilin@cuhk.edu.hk (H. Lin).

Present addresses:

<sup>a</sup>School of Geography and Environment, Jiangxi Normal University, No. 99, Ziyang Rd., Nanchang 330022, Jiangxi, China

<sup>b</sup>Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. 11A Datun Rd., Chaoyang District, Beijing 100101, China

<sup>c</sup>Wayz AI Technology Co. Ltd., No. 58, Xiangke Rd., Pudong District, Shanghai 201210, China

<sup>d</sup>Nanchang Center for Disease Control and Prevention, No 833, Lijing Rd, Honggutan New District, Nanchang 330038, Jiangxi, China

### Highlights

- LBS-requesting mobile device data is helpful to identify the generations of the COVID-19 epidemic spread at the early stage.
- A third-generation epidemic spread had already appeared in Wuhan from January 17 to 20, 2020, before the large-scale outbreak in mainland China started.
- The third-generation spread in Hubei province excluding Wuhan started from January 23 to 24, and the mixed first-generation and second-generation spreads strongly interacted to enhance the epidemic.
- The trend of the interactions between the first-generation and second-generation spreads was evident in the provinces adjacent to Hubei.
- The mixed first-generation and second-generation spreads appeared in most cities with confirmed cases in mainland China at the beginning of the outbreak.

### Abstract

**Objectives:** The outbreak of atypical pneumonia caused by the novel coronavirus (COVID-19) has currently become a global concern. The generations of the epidemic spread are not well known, yet these are critical parameters to facilitate an understanding of the epidemic. A seafood wholesale market and Wuhan city, China, were recognized as the primary and secondary epidemic sources.

Human movements nationwide from the two epidemic sources revealed the characteristics of the first-generation and second-generation spreads of the COVID-19 epidemic, as well as the potential third-generation spread.

**Methods:** We used spatiotemporal data of COVID-19 cases in mainland China and two categories of location-based service (LBS) data of mobile devices from the primary and secondary epidemic sources to calculate Pearson correlation coefficient,  $r$ , and spatial stratified heterogeneity,  $q$ , statistics.

**Results:** Two categories of device trajectories had generally significant correlations and determinant powers of the epidemic spread. Both  $r$  and  $q$  statistics decreased with distance from the epidemic sources and their associations changed with time. At the beginning of the epidemic, the mixed first-generation and second-generation spreads appeared in most cities with confirmed cases. They strongly interacted to enhance the epidemic in Hubei province and the trend was also significant in the provinces adjacent to Hubei. The third-generation spread started in Wuhan from January 17 to 20, 2020, and in Hubei from January 23 to 24. No obvious third-generation spread was detected outside Hubei.

**Conclusions:** The findings provide important foundations to quantify the effect of human movement on epidemic spread and inform ongoing control strategies. The spatiotemporal association between the epidemic spread and human movements from the primary and secondary epidemic sources indicates a transfer from second to third generations of the infection. Urgent control measures include preventing the potential third-generation spread in mainland China, eliminating it in Hubei, and reducing the interaction influence of first-generation and second-generation spreads.

**Keywords:** COVID-19; generations; mobile device data;  $q$  statistic; mainland China

## Introduction

An outbreak of atypical pneumonia caused by the 2019 novel coronavirus (COVID-19) was recognized from middle January, 2020, in Wuhan city, China. The novel coronavirus that infects human was first reported in Wuhan, Hubei province, China, on December 31, 2019 (Zhu et al. 2020). Early confirmed cases were mainly linked to a seafood wholesale market in Wuhan (Li et al. 2020a; Zhu et al. 2020). Epidemiological studies indicate that the COVID-19 epidemic has a basic reproductive number between 2 and 3 (Li et al. 2020a; Wu et al. 2020), which is lower than the 2003 severe acute respiratory syndrome (SARS) (Lipsitch 2003; Riley et al. 2003). Wuhan is a main transportation hub in central China, several million travelers ventured outward from the epidemic outbreak source in the first half of January, 2020, due to annual Chinese (Lunar) New Year holiday migrations. The large-scale outbreak started on January 19 (the first confirmed case reported outside Hubei province). Although strict transportation screening measures were activated by many cities in the next 3–4 days, the epidemic rapidly spread nationwide in a week. Moreover, COVID-19 infections have been identified in other countries and the current epidemic has become a global concern (Cohen and Normile 2020; Holshue et al. 2020; Rothe et al. 2020; Wang et al. 2020). The World Health Organization (WHO) declared the COVID-19 outbreak as a public health emergency of international concern (PHEIC) on January 30 (WHO 2020b). There is evidence that the epidemic outbreak in China and elsewhere spread along the paths of travel from Wuhan (Li et al. 2020b), and

local outbreaks could appear in other major cities of China with time lags (Wu et al. 2020).

Massive human movements via railways and domestic/international airlines from Wuhan, and the timing of Chinese New Year, has enabled the virus to spread nationwide and worldwide (Peeri et al. 2020). Control measures (e.g., travel quarantine and restrictions) in Wuhan were effective to delay the overall epidemic progression in mainland China and reduce the international case importations (Chinazzi et al. 2020). The Huanan seafood wholesale market and Wuhan were recognized as the primary and secondary epidemic centers, respectively, and therefore, the movements of populations from the two sources influenced the generations of the COVID-19 epidemic in mainland China, especially during the very early epidemic stage before the transportation measures activated by Wuhan and other cities. The first-generation (primary) spread of the epidemic was in part reflected by the human movement from the primary source (i.e., the seafood market), and the second-generation (secondary) spread was reflected by that from the secondary center (i.e., Wuhan city). They varied and interacted by region and time during the early epidemic progression, and had the potential clues to identify the third-generation spreads in various regions, which are mainly caused by the local cases instead of the imported ones. Here, using location-based service (LBS) data of mobile devices, we analyzed the spatiotemporal association of the confirmed COVID-19 cases and human movements from the sources of the epidemic outbreak, and revealed the first, second and potential third generation spreads of the COVID-19 epidemic in mainland China.

## Methods

### Data sources

We collected spatiotemporal data of COVID-19 cases in mainland China from the daily bulletins of the National Health Commission of the People's Republic of China (NHC) and various Provincial/Municipal Health Commissions. Some publicly available news and media were utilized as supplemental data. The final epidemic dataset was comparatively verified through the public platform of the 2019-nCoV-infected pneumonia epidemic from the Chinese Center for Disease Control and Prevention (China CDC 2020a). The dataset of the COVID-19 cases includes the following fields: date (starting from January 10, 2020), province code/name, city code/name, and numbers of daily new suspected/confirmed cases. From the above dataset, we can generate the cumulative number of daily confirmed cases at a specific city  $s$  and until a given end date  $t$ , which is denoted by  $y_{s,t}$ .

The human movement of populations from two epidemic sources (the Huanan seafood wholesale market and Wuhan), were considered to be associated with the spatiotemporal epidemic spread. The datasets of LBS requests from mobile devices were provide by Wayz Inc., Shanghai, China. The device trace datasets cover over 80% mobile devices supported by the three telecommunication operators in China. The LBS-requesting statistics are implemented every two hours with high-resolution location information. The raw data indicate the individual trajectories of numerous mobile devices with high-resolution spatiotemporal information, and can be easily aggregated in a specific spatial scale and within a given time step. For a subpopulation from the epidemic center, we can aggregate the device trace data from the start date to a given end date  $t$ , and the corresponding cumulative number at a specific city  $s$  is denoted by  $x_{s,t}$ . Multiple LBS requests within a time step are only counted once by a same device. Private individual information was deleted from the raw data of the mobile devices, and in this study, the device trace data was aggregated to the administrative cities and the epidemic date, i.e., the mobile device traces were associated with the

epidemic dataset according to date and location.

These aggregated statistics of mobile device traces are expected to be representative of the human migrations from the epidemic sources. Two epidemic sources were considered, including the seafood wholesale market and Wuhan city. The devices which activated their LBS requests in the market in November 2019 indicated the potential first-generation cases of the COVID-19 epidemic. And the potential second-generation cases were those which were activated in Wuhan in December 2019 and then traveled to other regions in January 2020.  $x_{s,t}^{(m)}$  and  $x_{s,t}^{(w)}$  are used to denote the spatiotemporal trajectories of the above two subpopulations of mobile devices, respectively. All the processing and aggregation of mobile device trace data were implemented by the provider. The final datasets include the daily counts of two categories of trajectories in all the administrative cities in mainland China. The cumulatively summed device traces had a spatially distributed consistency with the population distribution in mainland China (Figure 1). Two categories of trajectories mainly spread to the provinces adjacent to Hubei and several developed areas a longer distance from Hubei, such as Guangdong province, Zhejiang province and Beijing.

### Statistical Analyses

We considered the spread of the epidemic from the source in various space and time domains, and the corresponding associations with human movements were analyzed in several temporal divisions and spatial scales. Seven areas were delineated, including *i*) Wuhan city, *ii*) Hubei province excluding Wuhan, *iii*) Hubei province, *iv*) Hubei's adjacent provinces (Anhui, Chongqing city, Henan, Hunan, Jiangxi and Shaanxi), *v*) mainland China excluding Hubei, *vi*) mainland China excluding Wuhan, and *vii*) mainland China. Date periods were generated using three key date stamps, including January 10, 2020 (when the first 41 confirmed cases were reported in Wuhan), January 19 (when the large-scale outbreak started) and January 26 (the end of the first week of the large-scale outbreak).

Based on the above datasets of COVID-19 cases in mainland China and two categories of location-based service data of mobile devices from the epidemic sources, we calculated their Pearson correlation coefficient,  $r$ , and spatial stratified heterogeneity (SSH),  $q$ , statistics. Pearson correlation is usually used to evaluate the linear association between two variables and calculated as follows:

$$r_{xy} = \frac{\sum_{s=1}^n (x_{s,t} - \bar{x})(y_{s,t} - \bar{y})}{\sqrt{\sum_{s=1}^n (x_{s,t} - \bar{x})^2} \sqrt{\sum_{s=1}^n (y_{s,t} - \bar{y})^2}} \quad (1)$$

where  $r_{xy}$  denotes the correlation coefficient of COVID-19 spatiotemporal spread and human migrations from the epidemic source, within the period from the start date to a given end date  $t$ .  $y_{s,t}$  is the cumulative number of daily confirmed cases at city  $s$  and  $x_{s,t}$  is the cumulative number of device trajectories from the epidemic source, with the mean values of  $\bar{x}$  and  $\bar{y}$ , respectively.  $n$  is the number of the administrative cities in mainland China. In this study, we calculated two Pearson correlations with the spatiotemporal data of two categories of trajectories,  $x_{s,t}^{(m)}$  and  $x_{s,t}^{(w)}$ , to explore the associations between the epidemic spread and the human migrations from the seafood market and Wuhan, respectively.

**The GeoDetector  $q$  statistic is generally applied to quantitatively evaluate the SSH of an explained**

variable (Wang et al. 2010, 2016), and assess the determinant power of explanatory variables and their interaction, without linear assumptions (Yin et al. 2019). The fundamental formula of the  $q$  statistic is given by:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (2)$$

where  $q$  is the determinant power of the factor to the objective.  $N$  is the number of objective variable observations and  $\sigma^2$  indicates the variance of all the observations. The objective is stratified into  $L$  strata, denoted by  $h=1, 2, \dots, L$ , which is determined by the determinant factor.  $N_h$  is the number of observations and  $\sigma_h^2$  is the corresponding variance within stratum  $h$ . The value of  $q$  ranges from 0 to 1.

We calculate  $q$  statistic to assess the determinant power of human migrations from the epidemic source to COVID-19 spatiotemporal spread. Similarly, the spatiotemporal data of two categories of trajectories can be applied to calculate two  $q$  statistics for the two epidemic sources. Within the period from the start date to a given end date  $t$ , we implemented the stratification by the equal-interval division after ordering the trajectory data,  $x_{s,t}$ , and divided all the observations into 5 strata to calculate the  $q$  statistic of the cumulative trajectories,  $x_{s,t}$ , to the cumulative cases,  $y_{s,t}$ . This is a common stratification way to deal with the numerical independent variables (Yin et al. 2019), which can reduce the subjective influence of various stratifications to  $q$  statistics.

Moreover, for two or more determinant factors, an interaction  $q$  statistic can be calculated to measure their interaction influences (e.g., are they independent, or do they weaken/enhance each other?) (Wang et al. 2010). In this study, two categories of trajectories,  $x_{s,t}^{(m)}$  and  $x_{s,t}^{(w)}$ , were used to implement the stratifications and the corresponding  $q$  statistics were calculated, respectively, which are denoted by  $q^{(m)}$  and  $q^{(w)}$ . While the stratification was generated by the intersection between the above two individual stratifications, an interaction  $q$  statistic,  $q^{(m \cap w)}$ , can be calculated, where the symbol “ $\cap$ ” denotes the intersection between two strata layers. Various interaction types can be defined according to the comparison between  $q^{(m)}$ ,  $q^{(w)}$  and  $q^{(m \cap w)}$  (Wang et al. 2010). For instance, “ $q^{(m \cap w)} > q^{(m)}$  and  $q^{(w)}$ ” indicates a *bi-enhancement* interaction between two categories of trajectories in facilitating the spread of the epidemic (see Wang et al. 2010 for more details about the interaction  $q$  statistic).

Analyses in this study were performed with the use of the R software package (R Foundation for Statistical Computing) and thematic mapping was implemented in the ArcGIS platform (ESRI).

## Results

Similar to the spatial distributions of the mobile device traces (Figure 1), the Pearson correlations  $r$  and  $q$  statistics between the cumulatively summed cases and two categories of trajectories up to January 26, 2020 had a spatially distributed consistency with the population distribution among the administrative cities in mainland China (Figure 2). Two categories of trajectories had generally significant correlations and determinant powers of the epidemic spread, and both  $r$  and  $q$  decreased in distance from the epidemic sources. The first-generation and second-generation transmissions of the infection simultaneously appeared in many cities at the early stage of the outbreak. Specifically, devices activated in the market displayed higher values of  $r$  and  $q$  in several small and medium cities than devices activated in Wuhan city (Figures 2A and 2C). It is clear that many cities executed a quick response and activated transportation control measures, which helped control the first-

generation epidemic spreads. The  $r$  and  $q$  statistics of the devices activated in Wuhan, however, indicate that the second-generation spread still influenced many cities in the first week of the outbreak (Figures 2B, 2D and Table 1).

The market trajectories received a much higher Pearson correlation value to confirmed cases in Wuhan ( $r=0.6160$ ,  $p<0.001$ ) than Hubei province excluding Wuhan ( $r=0.3741$ ,  $p<0.001$ ) and mainland China excluding Hubei ( $r=0.3319$ ,  $p<0.001$ ). The correlations of Wuhan trajectories were 0.7438, 0.5874 and 0.5183 in the above three areas, respectively. The temporal correlation curves of both market and Wuhan trajectories have obvious decreasing trends from January 17 to 20, 2020 in Wuhan (Figure 3A), which indicates the potential start date of the third-generation epidemic spread. One week after this, market trajectories had higher Pearson correlation values than Wuhan trajectories, and the first-generation spread still had a serious influence in Wuhan (Figure 3A). Similarly, in Hubei province excluding Wuhan, the potential start date of the third-generation spread was from January 23 to 24 (Figure 3B). Moreover, the second-generation spread played a dominant role in the areas outside Wuhan, especially in Hubei province excluding Wuhan and the provinces adjacent to Hubei, since Wuhan trajectories had much higher values of correlations (Figures 3B and 3C). We found no obvious turning dates in the areas outside Hubei (Figures 3C and 3D), and the potential third-generation spread remains to be determined. The curves have remained stationary since January 22 in mainland China excluding Hubei (Figure 3D). The transportation control measures activated by many cities since January 21 appeared to have been successful in partially controlling the first-generation and second-generation epidemic spreads outside Hubei province.

We focused on the first week of the large-scale outbreak and calculated the  $q$  statistics of the two device-activation categories in introducing cumulative confirmed cases in various areas (Table 1). The determinant powers of both categories were extremely high and consistent in Wuhan ( $q=0.8909$ ,  $p<0.05$ ). Their temporal curves had the obvious decreasing trends from January 17 to 20 (Figure 4A), which validated the start date of the third-generation spread in Wuhan. Similar validation was observed in Hubei province excluding Wuhan (Figure 4B). Two categories of trajectories can explain nearly 100% SSH of the epidemic spread in Wuhan before the large-scale outbreak and the SSH increased constantly since the third-generation spread stage (Figure 4A). The market and Wuhan trajectories had close determinant powers in introducing the epidemic spread in Hubei province ( $q=0.4153$ ,  $q=0.4261$ , respectively, and  $p<0.001$ ). The  $q$  statistics reported that these two categories explained 41.53% and 42.61% SSH of the confirmed cases in Hubei. The determinant powers of the epidemic spread in Hubei province excluding Wuhan were 0.2084 ( $p<0.01$ ) and 0.2513 ( $p<0.001$ ), respectively. The  $q$  statistic values decreased in distance outside Wuhan or Hubei and showed that the determinant powers in mainland China excluding Hubei were 0.1610 ( $p<0.001$ ) and 0.1723 ( $p<0.001$ ), respectively.

In the first week of the outbreak, Wuhan trajectories received higher values of  $q$  statistics than market trajectories in Hubei province excluding Wuhan and in provinces bordering Hubei (Figures 4B and 4C). The second-generation spread contributed more influence in the areas surrounding the epidemic source. However, both two categories had close  $q$  statistic values in mainland China excluding Hubei (Figure 4D). The epidemic outside Hubei province appeared as a balanced pattern of mixed first-generation and second-generation spreads. Furthermore, the  $q$  statistics increased constantly outside Hubei province, indicating the increasing SSH of the epidemic spread (Figures 4C and 4D). More attention should be given to control of the trend of second-generation spread and to eliminate potential third-generation spread.

Taking into consideration of the interaction influences of two categories of trajectories, the interaction  $q$  statistics were calculated in various areas (Table 1). All the interaction types were *bi-enhancement* which indicates that two determinant factors (i.e., two categories of trajectories originated from two epidemic sources) enhance each other (the interaction  $q$  statistic is higher than each single  $q$  statistic but lower than the sum of two single  $q$  statistics). The determinant powers and interactions of two categories of trajectories in introducing the epidemic spread decreased in distance from the source to the rest of the nation. The interaction  $q$  statistic was 0.1925 (compared to the single  $q$  statistics of 0.1610 and 0.1723) in mainland China excluding Hubei. The interaction  $q$  statistic was 0.0786 (compared to the single  $q$  statistics of 0.0657 and 0.0642) in mainland China. Although the interaction strength was weak, the combination of both trajectory categories still carried more information about the spread of the epidemic throughout the country. The interaction  $q$  statistic of two categories of trajectories in Hubei province excluding Wuhan was 0.4063, which was close to the sum of two single  $q$  statistics (0.2084 and 0.2513) and much higher than each one individually. This interaction indicates strong *bi-enhancement* in facilitating the spread of the epidemic. Two categories of trajectories could significantly enhance each other to explain the SSH of the epidemic spread from Wuhan to other areas in Hubei province.

## Discussion

The majority of the earliest cases of the COVID-19 atypical pneumonia were linked to the seafood wholesale market in Wuhan, which is the most severely-affected city of the COVID-19 outbreak. The movements of populations from these two epidemic sources provided potential first-generation and second-generation spreads nationwide and worldwide. Here, based on LBS-requesting mobile device traces and spatiotemporal confirmed COVID-19 case data, we applied Pearson correlation and GeoDetector  $q$  statistics to analyze the spatiotemporal association between the confirmed cases' dynamic and human movements. Our findings provide important foundations to quantify the effect of human movement on the epidemic spread, to judge the epidemic generations, and to inform ongoing and future control strategies.

We concentrated on two datasets of LBS-requesting mobile devices associated with two sources linked to the first-generation and second-generation spreads provincewide and nationwide. Their traces were aggregated by date in administrative cities and linked to the spatiotemporal confirmed cases. It is notable that the COVID-19 outbreak had a strong consistency with human migrations from the epidemic sources. The confirmed cases had a clear linear correlation with two categories of trajectories from the sources to the rest of the nation. Moreover, both trajectory categories could generally indicate the epidemic spread in Hubei province and explain to a certain extent the SSH of the spread from Wuhan to the rest of Hubei province and throughout the rest of China. Our analyses provide a new perspective to explore the spread of the epidemics linked to human movement.

During the first week of the large-scale outbreak, the epidemic spread showed a spatially distributed consistency with the population distribution in mainland China. The majority of cities with confirmed cases had a mixed pattern of first-generation and second-generation spreads at the very beginning of the outbreak. Many cities activated quick response within 3–4 days and achieved efficient results in inhibiting the first-generation spread outside Hubei province. However, it still had a significant impact in Hubei province, especially playing the dominant role inside Wuhan city. Furthermore, among the other cities in Hubei province, the first-generation and second-generation spreads enhanced each other with a much higher interaction  $q$  statistic. This might be another signal

to identify the potential start date of the third-generation spread in a specific area. Due to the quick response and strict control measures in many cities, the interaction enhancement of the first-generation and second-generation spreads had a weak strength outside Hubei province. There is no evidence that any third-generation spread appeared outside Hubei in mainland China in the first week of the outbreak. Nevertheless, Hubei's adjacent provinces require more effective control measures, since the first-generation and second-generation spreads had an increasing trend.

Our analyses determined an appropriate approach to explore the spatiotemporal association between the epidemic transmission and human movement. Two categories of LBS-requesting mobile devices were used in this study to identify the potential close contacts to the primary and secondary epidemic sources. The datasets covered most devices with LBS requests in the given region and time period. However, the linkage between mobile devices and populations could be subject to information loss (e.g., users may replace their mobile devices with new ones). It is also extremely difficult to cover 100% potential close contacts in our datasets. The close contacts of these two populations while traveling before/after the outbreak were not collected, and therefore we cannot estimate the potential third-generation cases and their movements. This limitation involves future work with more universal-source data and high-performance computing capabilities. The COVID-19 epidemic data were collected through publicly available sources, and we processed the data of confirmed cases and device traces in the spatial scale of cities. Small-scale analyses could be more helpful to construct epidemic control programs in counties or communities within a city.

The spatiotemporal association between the spread of the epidemic and human movements indicates a transfer from second to third generations of the infection. This approach has made it possible to assess the start date of the third-generation spreads of COVID-19 epidemic and the interactions between first-generation and second-generation spreads across various regions all over the country. The proposed technique incorporating location-based service data of mobile devices can help identify the spatiotemporal generations at the early stage of the COVID-19 epidemic. It can be easily implemented and extended to the early exploratory study of other epidemics similar to COVID-19. The results indicate the spatiotemporal characteristics of the epidemic spread associated to human movements from epidemic sources and the potential spatiotemporal risks at the early stage of the outbreak. Control measures varying by location and time could be executed in different levels for various regions. For instance, cities with obvious third-generation spread require the strictest controls on both the exportations and the inside quarantine, cities should pay more attention to the importations and the inside quarantine if the first-generation and second-generation spreads have the strong interactive enhancements, and other cities require to focus on the control of the importations.

In conclusion, we found that the third-generation spread of the COVID-19 outbreak probably started during January 17 to 20, 2020 in Wuhan, the potential start date of the third-generation spread in Hubei province excluding Wuhan was from January 23 to 24, and the mixed first-generation and second-generation spreads strongly interacted to enhance the epidemic. The trend of the interactions between the first-generation and second-generation spreads was significant in the provinces adjacent to Hubei. The associations between the epidemic spread decreased with distance and had different temporal patterns from the epidemic sources, implying the potential epidemic generation-to-generation evolution on regional spatial scales. At the very beginning of the outbreak, the mixed first-generation and second-generation spreads appeared in most cities with confirmed cases. No obvious third-generation spread was detected outside Hubei province. The strict transportation

measures implemented in many cities appeared to have been effective in preventing any third-generation spread nationwide. The urgent control measures in Hubei province include weakening the third-generation spread and the interaction influence of the first-generation and second-generation spreads. Even with strict control strategies, effective measures to reduce transmission in the community are still required (Li et al. 2020a). A large increase in migration due to people returning from travel after the New Year holiday also introduces challenges to epidemic control (Wang and Zhang 2020). We recommend the urgent control measures of preventing potential third-generation spread in mainland China, eliminating it in Hubei, and reducing the interaction influence of first-generation and second-generation spreads.

#### **Ethics approval and consent to participate**

No individual data was collected and the ethical approval or individual consent was not applicable.

#### **Availability of data and materials**

The LBS-requesting mobile device data were provided by Wayz Inc., Shanghai, China and are not available for distribution due to the constraint in the consent. The dataset of the COVID-19 cases is available from multiple public sources.

#### **Funding**

This work was supported by the National Natural Science Foundation of China (41531179), the National Science and Technology Major Project of China (2016YFC1302504) and the Science and Technology Major Project of Jiangxi Province, China (2020YBBGW0007).

#### **Disclaimer**

The funders had no role in study design and conduct; data collection, management, analysis and interpretation; manuscript preparation, writing and review; decision to submit the manuscript for publication.

#### **Authors' contributions**

Conceptualization: Hui Lin, Jinfeng Wang and Bisong Hu; Resources and Data curation: Vincent Tao, Jingyu Qiu and Hui Lin; Formal analysis: Bisong Hu and Haiying Chen; Methodology: Bisong Hu and Jinfeng Wang; Software: Jinfeng Wang; Visualization: Bisong Hu; Validation, Bisong Hu and Jinfeng Wang; Writing—original draft: Bisong Hu; Writing—review and editing: Bisong Hu, Jinfeng Wang and Hui Lin.

#### **Conflict of interests**

We declare no competing interests.

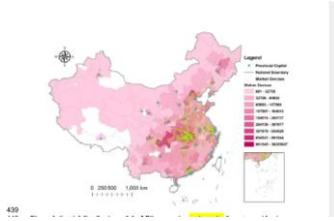
#### **Acknowledgments**

We thank Dr. Adam Thomas Devlin at the School of Geography and Environment, Jiangxi Normal University for the assistance in the proofreading work for the manuscript.

#### **References**

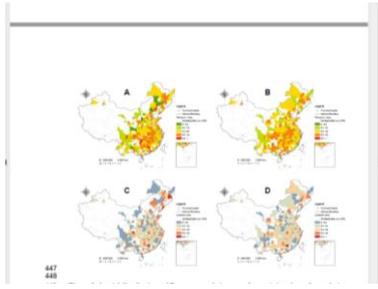
- China CDC. Public platform of the 2019-nCoV-infected pneumonia epidemic (in Chinese) [Internet]. 2020a [cited 2020 Feb 9]. Available from: <http://2019ncov.chinacdc.cn/2019-nCoV/>
- Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. 2020 Mar 6;eaba9757.
- Cohen J, Normile D. New SARS-like virus in China triggers alarm. *Science*. 2020 Jan 17;367(6475):234–5.
- Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*. 2020 Jan 31;
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med*. 2020a Jan 29;
- Li X, Zai J, Wang X, Li Y. Potential of large ‘first generation’ human-to-human transmission of 2019-nCoV. *J Med Virol*. 2020b Jan 30;
- Lipsitch M. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science*. 2003 Jun 20;300(5627):1966–70.
- Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int J Epidemiol*. 2020 Feb 22;dyaa033.
- Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, et al. Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science*. 2003 Jun 20;300(5627):1961–6.
- Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N Engl J Med*. 2020 Jan 30;
- Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The Lancet*. 2020 Jan;
- Wang F-S, Zhang C. What to do next to control the 2019-nCoV epidemic? *The Lancet*. 2020 Feb;395(10222):391–3.
- Wang J-F, Li X-H, Christakos G, Liao Y-L, Zhang T, Gu X, et al. Geographical Detectors- Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *Int J Geogr Inf Sci*. 2010 Jan 1;24(1):107–27.
- Wang J-F, Zhang T-L, Fu B-J. A measure of spatial stratified heterogeneity. *Ecol Indic*. 2016 Aug 1;67:250–6.
- WHO. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) [Internet]. 2020b [cited 2020 Feb 8]. Available from: [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-\(2019-ncov\)](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))
- Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*. 2020 Jan;
- Yin Q, Wang J, Ren Z, Li J, Guo Y. Mapping the increased minimum mortality temperatures in the context of global climate change. *Nat Commun*. 2019 Oct 11;10(1):1–8.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020 Jan 24;

Journal Pre-proof



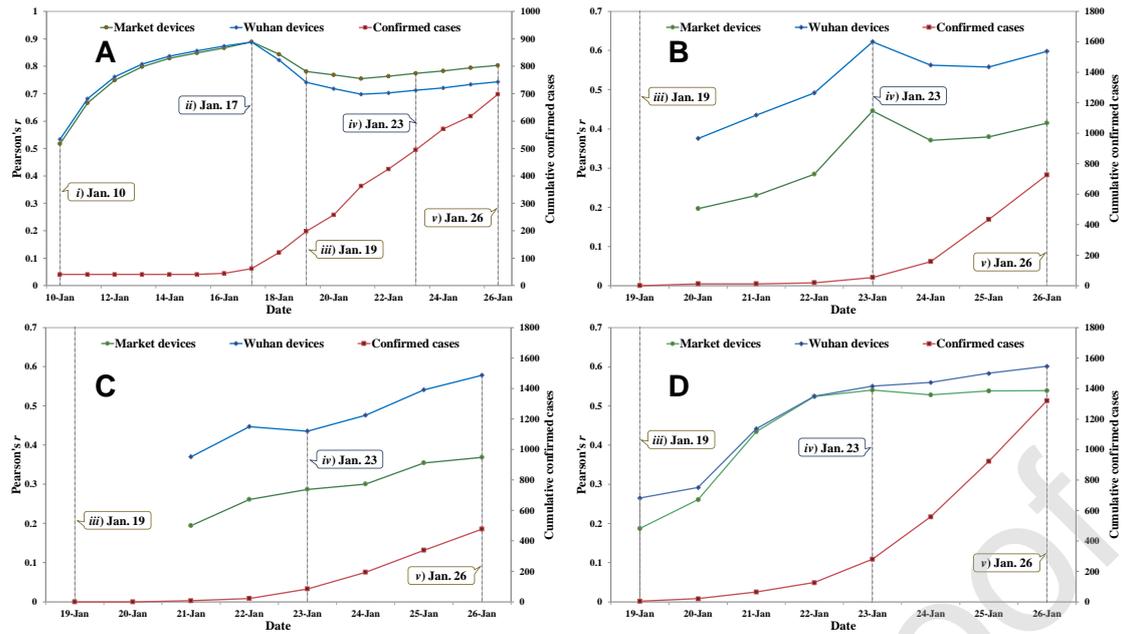
**Figure 1. Spatial distributions of the LBS-requesting trajectories from two epidemic sources.** Wuhan trajectories from January 1 to 27, 2020 are cumulatively summed and demonstrated as the thematic background among the administrative cities in mainland China. Market trajectories from November 7, 2019 to January 26, 2020, are illustrated as green-dotted point features. Note that the same device with multiple LBS requests within one day was only counted once.

Journal Pre-proof

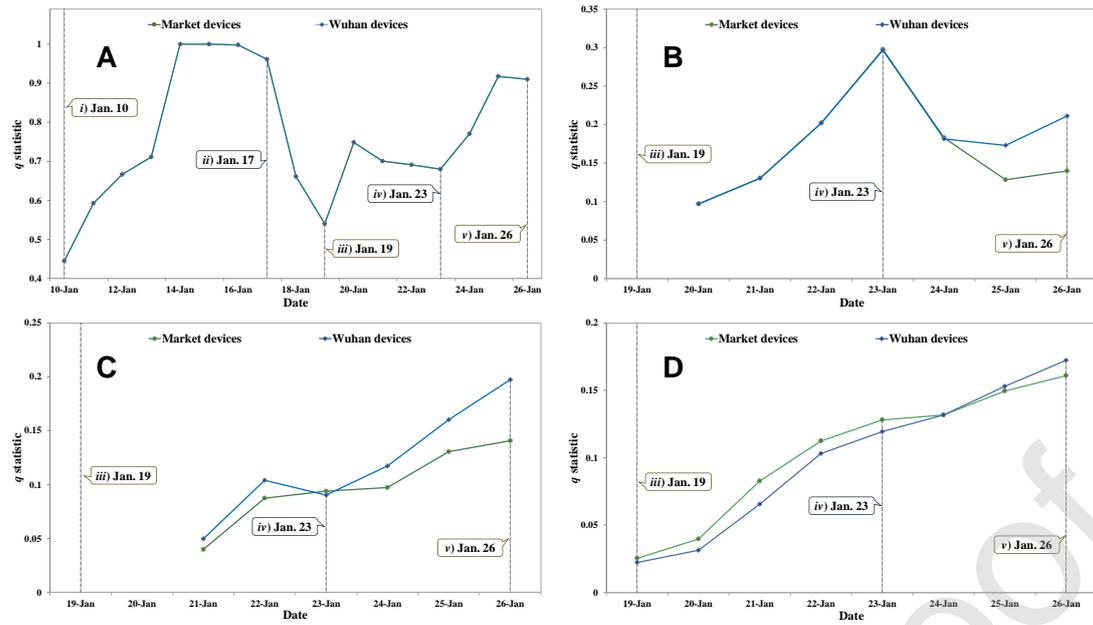


**Figure 2. Spatial distributions of Pearson correlations  $r$  and  $q$  statistic values of cumulative confirmed cases and LBS-requesting trajectories.** (A)  $r$  values of market trajectories. (B)  $r$  values of Wuhan trajectories. (C)  $q$  values of market trajectories. (D)  $q$  values of Wuhan trajectories. Not applicable or statistically insignificant ( $p>0.05$ ) values are shown in blank among the administrative cities in mainland China.

Journal Pre-proof



**Figure 3. Pearson correlations of cumulative confirmed cases, market trajectories and Wuhan trajectories until January 26, 2020.** (A) Wuhan city, (B) Hubei province excluding Wuhan, (C) Hubei's adjacent provinces, and (D) mainland China excluding Hubei. Note that the coordinate ranges in (A) differ from other subfigures. Several date stamps are emphasized, including: *i*) January 10, 2020 (when the first 41 confirmed cases were reported in Wuhan); *ii*) January 17 (when the third-generation spread started in Wuhan); *iii*) January 19 (when the large-scale outbreak nationwide started); *iv*) January 23 (when Wuhan implemented a full lockdown and the third-generation spread started in Hubei); and *v*) January 26 (the end of the first week of the outbreak).



**Figure 4.  $q$ -statistic values of market trajectories and Wuhan trajectories in introducing cumulative confirmed cases until January 26, 2020. (A) Wuhan city, (B) Hubei province excluding Wuhan, (C) Hubei's adjacent provinces, and (D) mainland China excluding Hubei. Note that the coordinate ranges in these subfigures differ from each other. Emphasized date stamps have the same interpretations as Figure 3.**

**Table 1. Pearson correlations  $r$  and  $q$  statistic values of cumulative confirmed cases, market trajectories and Wuhan trajectories in various areas of mainland China.** Pearson correlations were calculated until January 26, 2020, and  $q$  statistic values and interactions of/between two categories of trajectories in introducing cumulative confirmed cases were calculated from January 19 to January 26, 2020.

	Pearson correlations		GeoDetetor $q$ statistic				
	Market trajectories	Wuhan trajectories	Market trajectories	Wuhan trajectories	Interaction	Type	Strength
Wuhan city	0.6160 <sup>***</sup>	0.7438 <sup>***</sup>	0.8909 <sup>*</sup>	0.8909 <sup>*</sup>	-	-	-
Hubei Provence excluding Wuhan	0.3741 <sup>***</sup>	0.5874 <sup>***</sup>	0.2084 <sup>**</sup>	0.2513 <sup>***</sup>	0.4063	Enhance, bi-	Strong
Hubei Provence	0.6126 <sup>***</sup>	0.7637 <sup>***</sup>	0.4153 <sup>***</sup>	0.4261 <sup>***</sup>	0.5041	Enhance, bi-	Weak
Mainland China excluding Hubei	0.3319 <sup>***</sup>	0.5183 <sup>***</sup>	0.1610 <sup>***</sup>	0.1723 <sup>***</sup>	0.1925	Enhance, bi-	Weak
Mainland China excluding Wuhan	0.3593 <sup>***</sup>	0.5427 <sup>***</sup>	0.1709 <sup>***</sup>	0.1831 <sup>***</sup>	0.2096	Enhance, bi-	Weak
Mainland China	0.6098 <sup>***</sup>	0.7634 <sup>***</sup>	0.0657 <sup>***</sup>	0.0642 <sup>***</sup>	0.0786	Enhance, bi-	Weak

<sup>\*</sup>, <sup>\*\*</sup> and <sup>\*\*\*</sup> denotes  $p < 0.05$ ,  $p < 0.01$  and  $p < 0.001$ , respectively.