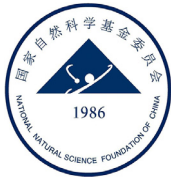




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Fundamental Research

journal homepage: <http://www.keaipublishing.com/en/journals/fundamental-research/>

Article

Assessing the spread risk of COVID-19 associated with multi-mode transportation networks in China

Xiao-Ke Xu^{a,1}, Xiao Fan Liu^{b,1}, Lin Wang^{c,1}, Ye Wu^{d,e}, Xin Lu^{f,g}, Xianwen Wang^h, Sen Pei^{i,*}^a College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China^b Web Mining Laboratory, Department of Media and Communication, City University of Hong Kong, Hong Kong SAR, China^c Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK^d School of Information Engineering, Putian University, Putian 351100, China^e School of Journalism and Communication, Beijing Normal University, Beijing 100875, China^f College of Systems Engineering, National University of Defense Technology, Changsha, China^g Department of Global Public Health, Karolinska Institute, Stockholm, Sweden^h WISE Lab, Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116085, Chinaⁱ Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

ARTICLE INFO

Article history:

Received 17 November 2021

Received in revised form 16 March 2022

Accepted 5 April 2022

Available online xxx

Keywords:

Human mobility

COVID-19

Transportation networks

ABSTRACT

The spatial spread of COVID-19 during early 2020 in China was primarily driven by outbound travelers leaving the epicenter, Wuhan, Hubei province. Existing studies focus on the influence of aggregated out-bound population flows originating from Wuhan; however, the impacts of different modes of transportation and the network structure of transportation systems on the early spread of COVID-19 in China are not well understood. Here, we assess the roles of the road, railway, and air transportation networks in driving the spatial spread of COVID-19 in China. We find that the short-range spread within Hubei province was dominated by ground traffic, notably, the railway transportation. In contrast, long-range spread to cities in other provinces was mediated by multiple factors, including a higher risk of case importation associated with air transportation and a larger outbreak size in hub cities located at the center of transportation networks. We further show that, although the dissemination of SARS-CoV-2 across countries and continents is determined by the worldwide air transportation network, the early geographic dispersal of COVID-19 within China is better predicted by the railway traffic. Given the recent emergence of multiple more transmissible variants of SARS-CoV-2, our findings can support a better assessment of the spread risk of those variants and improve future pandemic preparedness and responses.

1. Introduction

Emerging in late 2019, a novel coronavirus, SARS-CoV-2, spread rapidly from Wuhan to other cities in China in early 2020 [1]. By the end of February, the spatial spread of COVID-19, the disease caused by SARS-CoV-2, had been effectively curbed within China by strict national travel restrictions and home isolation policies enforced in late January 2020. Under strong Non-Pharmaceutical Interventions (NPIs), the epidemic went quickly under control in mainland China without massive local outbreaks outside the epicenter, Wuhan [2–4]. Unlike other countries with numerous unknown importations and subsequent prolonged community transmission [5], the single-source and short-term spread of COVID-19 in China provides a unique opportunity to study the impact of inter-city human mobility on the geographical spread of a novel pathogen [6,7].

Before Wuhan's lockdown on January 23, 2020, several billion trips took place during the Spring Festival season (i.e., “Chunyun”) and significantly accelerated the COVID-19 spread in China [8]. Indeed, out-bound mobility from Wuhan to other cities was found positively correlated with the number of confirmed cases in those cities [9]. However, previous studies have shown that epidemic spreading can be influenced by factors other than the sheer number of travelers from the epicenter. For instance, Balcan et al. demonstrated that different modes of transportation, e.g., road commuting and air traffic, may have different effects on local and long-range epidemic spread [10]. Belik et al. showed that recurrent daily commuting and diffusive human movement lead to different velocities of the propagating epidemic front [11]. Differential associations between the Gross Domestic Product (GDP) and COVID-19 cases were also reported in cities stratified by location, driven by different patterns of short- and long-range human mobility [12]. These prior works suggest that a precise assessment of epidemic spread risk requires

* Corresponding author.

E-mail address: sp3449@cumc.columbia.edu (S. Pei).¹ These authors contributed equally to this work.<https://doi.org/10.1016/j.fmre.2022.04.006>2667-3258/© 2022 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

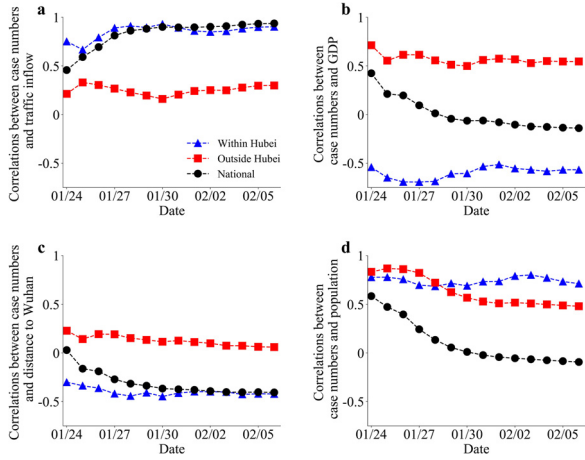


Fig. 1. The correlations between the cumulative numbers of reported COVID-19 cases and (a) traffic inflow from Wuhan, (b) GDP, (c) geographical distance to Wuhan, and (d) population size, for cities within and outside Hubei province, as well as all cities in China, from January 24 to February 6, 2020.

analysis of more detailed mobility data beyond the aggregated population flows.

A better understanding of the risk of COVID-19 spread associated with human mobility can guide future epidemic prevention and control. Considering the essential role of transportation in disease transmission, we characterize the impact of travels across the country focusing on the mode of transportation and from a network perspective. Specifically, we stratify destination cities by their provincial divisions, classify inter-city travels according to the means of transportation, and construct multi-mode transportation networks as the underlying structure that facilitates the transmission of SARS-CoV-2. We then analyze the transportation network structure and epidemic characteristics at different spatial scales, with the attempt to answer the following questions: (1) to what extent can different modes of transportation explain the number of COVID-19 cases? (2) Are different modes of transportation associated with varying levels of disease importation hazard? (3) How does the structure of the transportation network affect the progression of epidemic spread? We further use the effective distance, a network-based metric defined based on the transportation network structure, to predict the arrival time of COVID-19.

2. Materials and methods

2.1. Human mobility data

The inter-city human mobility in China is obtained from Baidu Qianxi (i.e., mobility) service and the Tencent Location-Based Services (LBS).

Baidu Qianxi service provides two relevant data for the inference of inter-city human mobility: a relative traffic volume index $O_i(t)$, which is a linearly scaled number of the daily traffic outflow from a city i , and the proportions p_{ij} of the traffic heading towards different destinations j . Therefore, the daily relative traffic volume from city i to j is $O_i(t)p_{ij}$. The traffic outflow from Wuhan that we used to plot Fig. 1 is calculated based on the Baidu Qianxi service by aggregating the daily relative traffic volume $O_i(t)p_{ij}$ from January 1 to January 23, 2020.

The traffic volume in the Baidu Qianxi service does not differentiate the mode of transportation. On the contrary, the Tencent LBS publishes traffic flows using the road, railway, and air transportation among 296 Chinese cities in 33 provinces from 321 2016 to 2019. Specifically, for each city, the top ten sources with the highest inbound traffic for each transportation mode and the top ten destinations with the highest outbound traffic are used to show transportation networks.

Since the Tencent LBS traffic data in 2020 were not available, we used their historical data to estimate human mobility before the lockdown. The lockdown in Wuhan was enforced on January 23, 2020, two days ahead of the Spring Festival. Before the Spring Festival holiday, a large number of residents in cities migrate to their hometown to reunion with families and celebrate the festival. Such massive population migration typically exhibits similar patterns during the two weeks prior to the Spring Festival each year. We used the traffic data between January 24 and February 16, 2018, which aligns to the two-week period before January 23, 2020, on the lunar calendar, to estimate human mobility during the study period. See Fig. S3 in Supplementary Information for a detailed plot showing the similarity between the 2018 Tencent LBS and 2020 Baidu Qianxi traffic. After the nationwide lockdown, we assume all inter-city mobility is stopped.

2.2. Imported cases from Wuhan

Health commissions in Chinese cities started publicly disclosing local case details online in January 2020. The public disclosure usually includes the demographic information (e.g., gender, age), travel history (e.g., departure and destination), and epidemiological history of cases (e.g., dates of symptom onset and case confirmation by epidemiological evidence). We collected all such online reports (more than 15,000 as of May 2021) and compiled a detailed line-list dataset which is updated on a bi-weekly basis [13]. Imported cases from Wuhan are defined as the reported individuals who have documented travel history with Wuhan as its departure or interim transfer city.

2.3. The construction of multi-mode transportation networks

The road, rail, air, and aggregated transportation networks $G = \{V, E\}$ were constructed at both city and province levels. Nodes $v \in V$ in the network represent cities or provinces, and directed edges $e \in E$ are time-stamped and weighted by the reported traffic volumes through road, railway, air, and all transportation modes aggregated between cities or provinces on each day.

2.4. PageRank centrality

Given a transportation network with directed edges weighted by the passenger flux between the nodes, the PageRank centrality PR_i [14] for node i is defined as

$$PR_i = \alpha \sum_{j, e_{ij} \in E} PR_j / L_j + (1 - \alpha) / N \quad (1)$$

where $\alpha = 0.85$ is a damping factor, j is a neighbor of node i , L_j is the out-degree of node j in the network, and N is the number of nodes in the network. The algorithm is realized by the *NetworkX* Python package.

2.5. Effective distance

Given a transportation network with directed edges weighted by the passenger flux between the nodes, the matrix P with $0 \leq P_{ji} \leq 1$ quantifies the fraction of the passenger flux with destination j emanating from node i , i.e., $P_{ji} = F_{ji} / F_i$, where $F_i = \sum_j F_{ji}$. The effective distance d_{ji} [15] from a node i to a connected node j is defined as

$$d_{ji} = 1 - \log P_{ji} \quad (2)$$

where d_{ji} ranges from 1 to infinity. $d_{ji} = 1$ means all the travelers from city i arrive in city j , and $d_{ji} \gg 1$ means that very few travelers from city i go to city j . d_{ji} is a directional measure and depicts travelers' preference departing from the source city i to other cities.

The concept of effective distance reflects the idea that a small fraction of traffic $i \rightarrow j$ is effectively equivalent to a large distance, and vice versa. On the basis of this, we can define the directed length $\lambda(\tau)$ of an ordered path $\tau = \{i_1, \dots, i_L\}$ as the sum of effective lengths along the

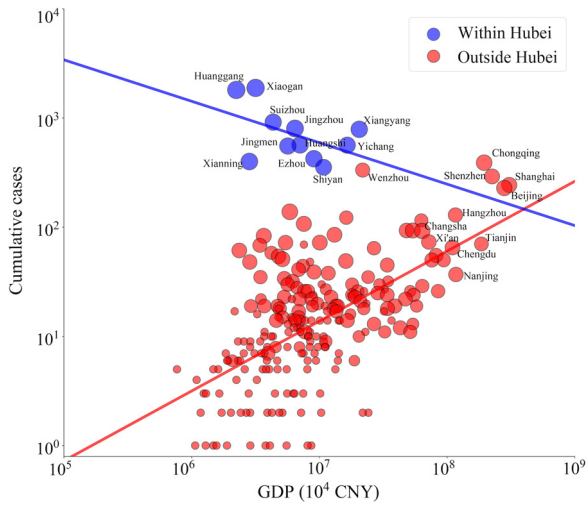


Fig. 2. Relationship between the cumulative confirmed cases on February 6, 2020 and cities' GDP in 2020. Cities within and outside Hubei province are highlighted in blue and red, respectively. Node sizes are proportional to the log-transformed average daily outflows from Wuhan. See Supplementary Information and Fig. S2 for the same plots between the cities' cumulative confirmed cases and traffic inflow from Wuhan, distances to Wuhan, and city population in 2020.

legs of the path. Moreover, we define the effective distance d_{ij} from an arbitrary reference node i to another node j in the network by the length of the shortest path from i to j :

$$D_{ji} = \min_{\tau} \lambda(\tau). \quad (3)$$

A short effective distance can accelerate virus spreading, while a large effective distance will hinder the spreading of diseases.

2.6. Data availability

The relevant COVID-19 data are collected from public sources and will be publicly available at GitHub upon publication.

3. Results

3.1. The Simpson's Paradox

The total traffic flow from Wuhan to other cities in China is highly correlated with the cumulative number of reported cases (case number hereafter) towards the end of the first wave in early February 2020, with a correlation coefficient of around 0.9 (black line, Fig. 1a) [6,9]. However, a closer look reveals a more complex pattern. When stratifying cities into two subgroups according to their provincial divisions, cities located outside Hubei province have a much lower correlation between case numbers and inbound traffic from Wuhan. Similarly, several geographic, demographic, and socioeconomic features also exhibit differential associations with case numbers. For instance, whereas case numbers are negatively correlated with the GDP for cities within Hubei province, the association becomes positive for cities outside Hubei province (Figs. 1b and 2). In China, cities with higher GDP are generally more populated and urbanized. The more frequent human interactions in urban settings could potentially increase the transmission risk of COVID-19. Cities with higher GDP are also likely to be transportation hubs with more human mobility from other locations. The association between GDP and COVID-19 cases reflects the indirect effect mediated by these factors impacting disease transmission. In addition, the correlations between case numbers and the geographical distance to Wuhan (Fig. 1c) and population size (Fig. 1d) also demonstrate differences for cities within and outside Hubei province. Similar phenomena were also observed if the estimated total number of infections (both

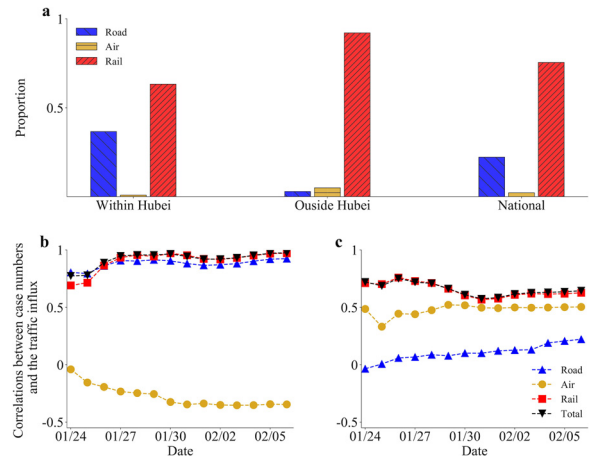


Fig. 3. (a) The proportion of road, rail, and air traffics departing from Wuhan to other cities. The correlations between the cumulative COVID-19 case numbers from January 24 to February 6, 2020 and the influx via road, air, and rail from Wuhan to other cities within (b) and outside (c) Hubei province before the lockdown.

documented and undocumented) were used [12,16], see Supplementary Information and Fig. S1 for details.

The above phenomena are attributed to a common statistical fallacy - the Simpson's Paradox [17,18]. In statistical analysis, merely assuming that individuals would align with the statistics of the whole group can wrongly characterize the peculiarities of individuals or subgroups. In our case, the association between case numbers and traffic inflow from Wuhan is significantly reduced when stratifying cities by their proximity to Wuhan. Other examined geographic, demographic, and socioeconomic features also show different explanatory powers to case numbers (Figs. 2 and S2). These findings indicate that lumping all cities together into the analysis of COVID-19 spread risk can overlook the subtle but important difference among subgroups of cities and hence undermine the precision of the risk assessment. As a result, an in-depth analysis is needed for better identification of risk factors that are more specific to individual cities.

3.2. Land traffic as the dominant driver of short-range epidemic spread

People choose different means of transportation to travel at various spatial scales. In China, road and air transportation are typically used for short-range and long-haul travels, respectively, and railway is used for both short- and long-range mobility. Among the outflow from Wuhan during the "Chunyun" period, rail traffic accounts for a dominating 75.5% of the total volume (Fig. 3a). For short-range trips within Hubei province, road transportation also contributes a significant proportion, with air transportation playing a minimal role. In contrast, air transportation is more frequently used than road transportation for traveling to destinations outside Hubei province.

Different transportation modes can shape the geographical spread of an emerging infectious disease with complex effects. Existing research on COVID-19 has drawn inconsistent conclusions. For instance, Zheng et al. [19] showed that COVID-19 case numbers are better correlated with bus and train traffics than air traffic, while Zhao et al. [20] found that case numbers are significantly associated with rail traffic but not with road and air traffics. Here, by stratifying destination cities into subgroups within and outside Hubei province, we found a mixing effect. For cities within Hubei province, case numbers are positively correlated with the number of travelers imported from Wuhan via road and railway before lockdown but are negatively correlated with air traffic (Fig. 3b). For cities outside Hubei province, however, case numbers show an increased correlation with air traffic and the correlation with road traffic disappears (Fig. 3c).

Table 1
Relative importation risks of different transportation modes for cities outside Hubei province.

	Number of imported cases	
	Model 1	Model 2
N	48	48
total arrival	0.005***	-
road arrival	-	0.007***
rail arrival	-	0.005***
air arrival	-	0.099***
R_2	0.119	0.212
F-value	90.60	59.81

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Intuitively, a high-volume transportation mode should primarily drive the spatial spread of a disease. This intuition apparently holds for cities within Hubei province, where land traffic and case numbers are strongly correlated. However, for cities distant from the epicenter, air traffic, the transportation mode that accounts for only a small proportion of the total volume, becomes a strong driver of COVID-19 spread. In the next two subsections, we explore two possible explanations for this counter-intuitive phenomenon: the importation risk associated with different transportation modes and the network effect of transportation hubs in long-distance travel.

3.3. Higher importation risk associated with air transportation

To compare the importation risk for different transportation modes, we constructed two linear regression models to explore the association between the numbers of imported cases in cities outside Hubei province and traffic flow from Wuhan. In the first model, we used the total number of passengers from Wuhan as the explanatory variable; in the second, the numbers of passengers using the road, rail, and air transportation were used as explanatory variables. In the analysis, 48 cities with information on imported cases identified through epidemiological investigation were included. As expected, case importation is positively correlated with the total influx from Wuhan (Table 1). However, after breaking down the total influx by the mode of transportation, we found a considerably higher regression coefficient for air traffic (0.099) than those for the road (0.007) and rail (0.005) transportation. This finding suggests that the importation risk associated with airplane passengers is much higher than that of travelers using land traffic. Collinearity test (see Section 2 in Supplementary Information) assures that the higher regression coefficient for air traffic is not caused by multicollinearity between the explanatory variables. Moreover, these results are robust when the total number of cases in each city is used as the dependent variable (see Table S1 for details). The reasons for the higher importation risk associated with air traffic need to be further explored.

3.4. Node centrality explains increased importation risk at transportation hubs

As a regional economic center, Wuhan is connected to its peripheral cities in a relatively simple star-like hub-and-spoke structure. However, when considering long-range disease spread, one has to consider transits between origins and destinations and the mobility network formed by interim transits and city-hopping travelers. The mobility flows in China can be described by a complex multiscale network spanning several orders of magnitude in intensity and spatial scales, as shown in Fig. 4. Note the connections indicate the origins and final destinations without showing transit stops. The Chinese road network appears as a grid-like lattice connecting neighboring subpopulations (Fig. 4a), whereas the air traffic network is composed of long-range connections (Fig. 4b). The rail network covers the middle ground between the two (Fig. 4c).

For a disease spreading from Wuhan to distant locations, there exist a number of possible paths through potential multimode transits. For instance, a traveler can take long-range flight (possibly with interim transits) to major air traffic hubs and then travel to peripheral cities through the road network. In another option, the traveler can reach distant destinations through multiple train transits or a hybrid train-road travel.

Hub nodes in the long-range transit networks may be exposed to a higher importation risk as they aggregate the traffics of transiting travelers from/to surrounding locations. Here, we quantify this risk using several network centrality measures, including in-degree (defined as the number of cities that have traffic flow to a city), in-strength (defined as the total traffic influx to a city), and PageRank [14], in the three transportation networks. The degree and strength are both local centrality measures, i.e., considering only the immediate neighbors of a node, while PageRank is a global centrality measure that resembles the graph Laplacians. Specifically, PageRank captures the effect of network structure by taking into account the importance of nodes' neighbors recursively. The correlations between the cumulative case numbers in a city and the network centrality measures, as well as the traffic influx from Wuhan, are shown in Fig. 4 for the three networks.

Node centrality measures in the road and air traffic networks better explain the number of COVID-19 cases found in a city than the total traffic influx from Wuhan (Fig. 4d and e). Notably, in the road network, the degree centrality can better explain the COVID-19 case numbers than the traffic influx from Wuhan and the global centrality measure, PageRank, especially at the early stage of disease spreading (Fig. 4d). This is possibly due to that road traffic is mostly involved in local and short-range epidemic spread. For the airline network, PageRank and strength have the highest correlation with case numbers and can also better explain the epidemic size than traffic influx from Wuhan (Fig. 4e). For the rail network, the traffic influx from Wuhan has similar explanatory power to case numbers as local network centrality measures (Fig. 4f).

This finding confirms that the railway arrivals, which account for the highest proportion of Wuhan's outbound traffic, play a dominant role in disease spreading. However, the total volume of road and air traffic from Wuhan cannot well explain case numbers in Chinese cities. For the lattice-like road network, high-degree nodes are usually regional centers, while for the structurally heterogeneous air traffic network, a higher in-strength or PageRank centrality indicates that the node is likely a transportation hub in the network. The better predictive power of strength and PageRank in the air traffic network and degree in the road network implies that, amid the short period during which COVID-19 spread from Wuhan to other cities, long-range disease spreading potentially passed through national transportation hubs and regional centers before arriving in cities that do not have a direct connection to Wuhan.

3.5. Traffic flux in the railway network better predicts the epidemic arrival time

Apart from case numbers, another key question on disease spreading is the arrival time at distant locations following an outbreak. The shortest effective distance proposed by Brockmann and Helbing [15] has shown predictive power in estimating epidemic arrival times in different countries using the global airline network. Refined effective distances that consider all possible transmission paths have also been proposed to predict the arrival times of an infectious disease [21,22]. The shortest effective distance between two nodes in a network is calculated using the path with the shortest point to point effective distance, defined as the inverse of the proportion of traffic influx from the source in the destination's total influx. A shorter effective distance means the destination accepts a larger proportion of travelers from the outbreak source.

We compute the shortest effective distance D_{eff} from Wuhan to other provinces in each transportation network. The epidemic arrival times T_{arr} , i.e., the date of first reported case in a province, is best correlated

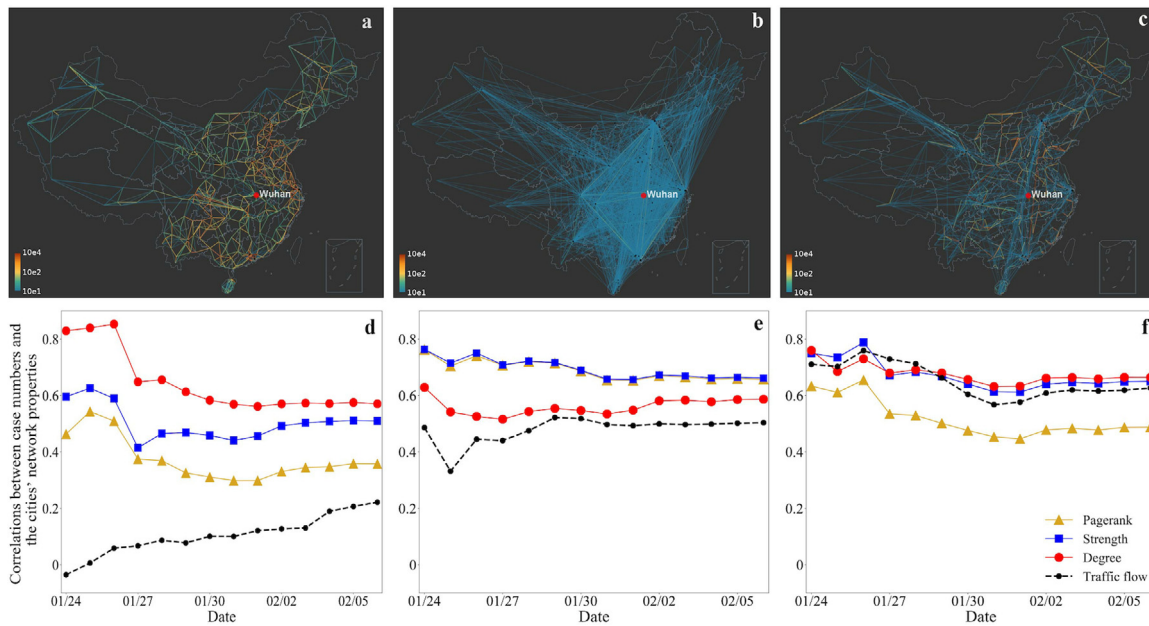


Fig. 4. Transportation networks in China. The (a) road, (b) air, and (c) rail transportation networks are formed by inter-city travels. The width and color (from blue to red) of the edges represent the mobility flow intensity on a logarithmic scale. Wuhan locates near the geographic center of China and is a hub in all three transportation networks. The correlations between case numbers and cities' network centrality measures (in-degree, in-strength, PageRank, and traffic influx from Wuhan) in the road, air, and rail transportation networks are shown in (d), (e), and (f), respectively.

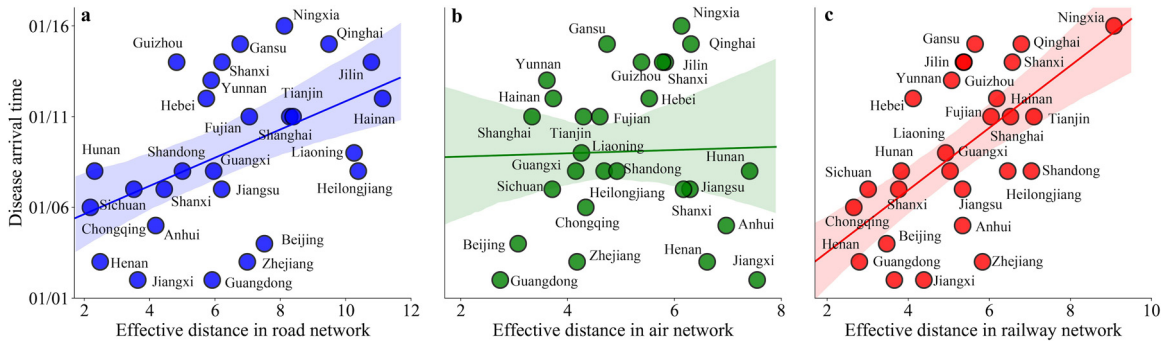


Fig. 5. The correlation between epidemic arrival times T_{arr} and effective distances D_{eff} in different transportation networks: (a) road ($r = 0.44$), (b) air ($r = 0.02$), (c) railway ($r = 0.60$).

with the D_{eff} in the rail network (with a correlation coefficient $r = 0.60$), and less correlated with the road network ($r = 0.44$), as shown in Fig. 5a and c. In contrast to the dominate role of the airline network in international epidemic spread, the correlation between T_{arr} and D_{eff} in the air traffic network is nominal ($r = 0.02$) for COVID-19 spread in China (Fig. 5b). The strong association between the shortest effective distance in the land transportation networks and epidemic arrival time shows that a destination with a larger proportion of arrivals from Wuhan may be exposed to the disease earlier. This analysis indicates that geographical scale matters in the assessment of the spread risk of an emerging infectious disease. While international disease spread is largely driven by long-range flights, progression of COVID-19 spread within China is determined by high-volume land transportation.

It is important to examine whether railways are also major drivers of SARS-CoV-2 transmissions in other provinces outside the epicenter. However, as case numbers were low in other provinces, it is not feasible to perform the same analysis for each individual province. As an alternative, we analyzed the mobility data in the top 10 provinces with the largest population. Specifically, we computed (1) the proportion of railway travels from the province capital to other cities within that province, and (2) the proportion of railway travels from the province

Table 2

The proportion of railway transportation from the capital city of each province to other cities within and outside the province. Analysis was performed for the top 10 provinces with the largest resident population.

Province (Capital City)	Within Province (%)	Outside Province (%)
Guangdong (Guangzhou)	52.1	86.3
Shandong (Jinan)	56.2	80.1
Henan (Zhengzhou)	55.8	89.1
Jiangsu (Nanjing)	61.1	72.8
Sichuan (Chengdu)	52.8	85.1
Hebei (Shijiazhuang)	52.7	79.9
Hunan (Changsha)	51.1	82.3
Zhejiang (Hangzhou)	51.7	77.0
Anhui (Hefei)	52.7	77.2
Hubei (Wuhan)	61.1	92.1

capital to cities outside that province. Results in Table 2 indicate that the railway is still the leading mode of inter-city transportation within these provinces, similar to Hubei province. As a result, railways were also potentially a major driver of intra-provincial transmission of SARS-CoV-2 in other provinces. We additionally checked whether the proportion of

railway travel from Wuhan to other cities increased during the Chunyun period. We found this proportion remained unchanged throughout the year (Fig. S4). These analyses imply that the identified spread risk associated with railways may be generalized beyond the epicenter Hubei and the Chunyun period.

4. Conclusion and discussion

In early 2020, COVID-19 swept across the globe in a span of just a few months. Understanding the spatial spread of COVID-19 is critical for the preparedness for future emerging infectious diseases. Previous studies found that the spread risk of COVID-19 is highly correlated with the number of travelers from the epicenter. Here, we take a closer look by stratifying population flows according to the mode of transportation and find more nuanced patterns. In particular, different transportation networks induce distinct spread risk of COVID-19 in other locations depending on their proximity to the epicenter and network structure. While disease spread in peripheral cities close to Wuhan is dominantly affected by road and train traffics, the long-range spread is impacted by a number of factors, including higher importation risk associated with air traffic and increased risk of transportation hubs that aggregate transiting travelers. The traffic flux in the railway network can better predict the arrival time of COVID-19 in China. A few limitations exist for this study. First, cases were mostly reported within Hubei province, which may impact the computation of associations in Figs. 1 and 2. Second, in this study, we stratified locations according to administrative boundaries (i.e., within and outside Hubei province). Analysis based on the distance to the epicenter should yield qualitatively similar results. However, determining the geographical scale of the distance separating the short- and long-range spread, which may depend on the catchment area of the epicenter, needs to be further studied.

The implication of our study has two folds. First, with vaccines being disseminated worldwide and economic activities expected to recover gradually, the risk of importation/reintroduction of COVID-19 through long-haul travels shall not be overlooked. Therefore, it is imperative to have an accurate assessment of the COVID-19 spread risk to prevent the re-emergence of the virus and introduction of more transmissible variants. Second, accurate estimation of the spread risk and the arrival time of emerging infectious diseases needs to consider the patterns of multi-scale human mobility and the associated traffic network. Analysis lumping all locations and aggregated mobility data together may lead to results that do not apply to a certain subgroup of locations.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [61773091 and 62173065 to X.-K. X., 11975025 to L. W., 11875005 to Y. W., 72025405 and 82041020 to X. L., 71974029 to X. W.]; the Grand Challenges ICODA pilot initiative, delivered by Health Data Research UK and funded by the Bill & Melinda Gates Foundation and the Minderoo Foundation [to X.F. L.]; US CDC Grant 20U01CK000592 [to S.P.]; and US CDC and CSTE Grant NU38OT00297 [to S.P.].

Supplementary materials

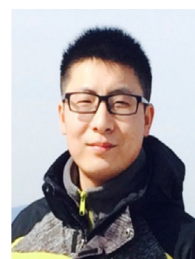
Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.fmre.2022.04.006](https://doi.org/10.1016/j.fmre.2022.04.006).

References

- [1] Q. Li, X. Guan, H. Li, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, *N. Engl. J. Med.* 382 (2020) 1199–1207.
- [2] M. Chinazzi, J. Davis, M. Aielli, et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, *Science* 368 (2020) 395–400.
- [3] M.U.G. Kraemer, C. Yang, B. Gutierrez, et al., The effect of human mobility and control measures on the COVID-19 epidemic in China, *Science* 368 (2020) 493–497.
- [4] S. Lai, N. Ruktanonchai, L. Zhou, et al., Effect of non-pharmaceutical interventions to contain COVID-19 in China, *Nature* 585 (2020) 410–413.
- [5] L. du Plessis, J.T. McCrone, A.E. Zarebski, et al., Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK, *Science* 371 (2021) 708–712.
- [6] H. Tian, Y. Liu, Y. Li, et al., An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China, *Science* 368 (2020) 638–642.
- [7] Z. Du, L. Wang, S. Cauchemez, et al., Risk for transportation of coronavirus disease from Wuhan to other cities in China, *Emerg. Infect. Dis.* 26 (2020) 1049–1052.
- [8] P. Zhong, S. Guo, T. Chen, Correlation between travellers departing from Wuhan before the Spring Festival and subsequent spread of COVID-19 to all provinces in China, *J. Travel Med.* 27 (2020) taaa036.
- [9] J.S. Jia, X. Lu, Y. Yuan, et al., Population flow drives spatio-temporal distribution of COVID-19 in China, *Nature* 582 (2020) 389–394.
- [10] D. Balcan, V. Colizza, B. Gonçalves, et al., Multiscale mobility networks and the spatial spreading of infectious diseases, *Proc. Natl. Acad. Sci. USA* 106 (2009) 21484–21489.
- [11] V. Belik, T. Geisel, D. Brockmann, Natural human mobility patterns and spatial spread of infectious diseases, *Phys. Rev. X* 1 (2011) 011001.
- [12] X.K. Xu, L. Wang, S. Pei, Multiscale mobility explains differential associations between the gross domestic product and COVID-19 transmission in Chinese cities, *J. Travel. Med.* 28 (2021) taaa236.
- [13] X.F. Liu, X.K. Xu, Y. Wu, Mobility, exposure, and epidemiological timelines of COVID-19 infections in China outside Hubei province, *Sci. Data* 8 (2021) 54 54–54.
- [14] L. Page, S. Brin, R. Motwani, et al., The Pagerank Citation Ranking: Bringing Order to the Web, Stanford InfoLab, 1999.
- [15] D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena, *Science* 342 (2013) 1337–1342.
- [16] R. Li, S. Pei, B. Chen, et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science* 368 (2020) 489–493.
- [17] C.R. Blyth, On Simpson's Paradox and the sure-thing principle, *J. Am. Stat. Assoc.* 67 (1972) 364–366.
- [18] M.A. Hernán, D. Clayton, N. Keiding, The Simpson's Paradox unraveled, *Int. J. Epidemiol.* 40 (2011) 780–785.
- [19] R. Zheng, Y. Xu, W. Wang, et al., Spatial transmission of COVID-19 via public and private transportation in China, *Travel Med. Infect. Dis.* 34 (2020) 101626 101626–101626.
- [20] S. Zhao, Z. Zhuang, J. Ran, et al., The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: a data-driven correlational report, *Travel Med. Infect. Dis.* 33 (2020) 101568 101568–101568.
- [21] F. Iannelli, A. Koher, D. Brockmann, et al., Effective distances for epidemics spreading on complex networks, *Phys. Rev. E* 95 (2017) 012313 012313–012313.
- [22] L. Wang, J.T. Wu, Characterizing the dynamics underlying global spread of epidemics, *Nat. Commun.* 9 (2018) 218.



Xiao-Ke Xu is a Full Professor at the College of Information and Communication Engineering, Dalian Minzu University, Dalian, China. He received his Ph.D. degree from the College of Information and Communication Engineering, Dalian Maritime University, China, in 2008. He was a Post-Doctoral Fellow with Hong Kong Polytechnic University and a Visiting Scholar with the City University of Hong Kong. His current research interests are in network community detection, transmission dynamics of infectious diseases, and data mining in social networks.



Sen Pei is an Assistant Professor in the Department of Environmental Health Sciences at Mailman School of Public Health, Columbia University. He received his Ph.D. in Mathematics from Beihang University in 2015. Before joining the faculty at Columbia in 2021, he worked as a postdoctoral research scientist and an associate research scientist at the Mailman School of Public Health. He studies environmental, social, and ecological determinants of infectious disease, with a particular interest in mathematical modeling, statistical inference, and real-time forecasting of seasonal and emerging infectious diseases. His recent studies focus on influenza, SARS-CoV-2, and antimicrobial-resistant pathogens.