The spatial dissemination of COVID-19 and associated socio-economic consequences

Yafei Zhang[1,2], **Lin Wang**[1], **Jonathan J. H. Zhu**[2,*], **and Xiaofan Wang**[1,3,*]

[1]Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China
[2]Department of Media and Communication, and School of Data Science, City University of Hong Kong, Hong Kong S.A.R., China
[3]Department of Automation, Shanghai University, Shanghai 200444, China
[*]Correspondence to: j.zhu@cityu.edu.hk, xfwang@sjtu.edu.cn

ABSTRACT

The ongoing coronavirus disease 2019 (COVID-19) pandemic has wreaked havoc worldwide with millions of lives claimed, human travel restricted, and economic development halted. Leveraging city-level mobility and case data across mainland China, our analysis shows that the spatial dissemination of COVID-19 in mainland China can be well explained by the human migration from Wuhan and there will be very different outcomes if the COVID-19 outbreak occurred in other cities. For example, the outbreak in Beijing or Guangzhou would result in a ∼90% increase of COVID-19 cases at the end of the Chinese New Year holiday. After the implementation of a series of control measures, human mobility had experienced substantial changes toward containing the spread of COVID-19. Our results also suggest an inequality of economic deprivation as less developed areas generally suffered more severe economic recession during the COVID-19. Intuitively, it's anticipated that cities with more confirmed cases would suffer more economic losses. However, for cities outside of Hubei province, we don't observe such a phenomenon. Our work has important implications for the mitigation of disease and the reevaluation of the social and economic consequences of COVID-19 on our society.

1 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which caused coronavirus disease 2019 (COVID-19), emerged in Wuhan (the provincial capital of Hubei province) in December 2019 and diffused quickly across mainland China, coinciding with mass human migration during the Spring Festival period[1,2]. Given the migration scale (estimated to be three billion trips over the festival period) and the position of Wuhan in the national transportation network, combating the dissemination of SARS-CoV-2 becomes urgent and very challenging.

In response to the rapid escalation of COVID-19 cases and deaths and to contain the threat of COVID-19, Wuhan, the epicenter of COVID-19 outbreak in mainland China, was put on lockdown on 23 January 2020, followed by the lockdown of other cities in Hubei province and national lockdown afterwards. Unprecedented measures were also implemented, including the isolation of individuals who had related symptoms, quarantine of people who returned from Wuhan, mandatory mask-wearing at public indoor venues, closure of non-essential businesses and activities, and extension of the Chinese New Year holiday period[2,3]. These bold acts, accompanying with other social distancing policies and improved clinical testing capacity, had drastically shifted the rapid evolution of COVID-19 outbreak in mainland China[2,4,5]. After the extended holiday, national reopening was then put in force orderly due to the notably positive trending of COVID-19 containment.

Human movements are thought to play a crucial role in shaping the spatio-temporal transmission of infectious diseases[6–14]. Using human mobility and case data covering more than 360 cities in mainland China, we assess how human mobility drives the spatial dissemination of COVID-19 across the country and what it would be if the COVID-19 outbreak happens in cities other than Wuhan. We find that the spatial dissemination of COVID-19 across mainland China can be well explained by the human flow from Wuhan and city population, which is quite different from classic disease spreading process in complex networks where cities located at central positions in the mobility network would get more infections due to the travel of infected

people. This also indicates the effectiveness of the implemented strict control measures in mainland China, where most of the infected people were quarantined or isolated during the lockdown, thus largely preventing further transmission to other areas. Although Wuhan is one of the hubs in the national transportation network, it would be much more dangerous if the COVID-19 outbreak happens in other cities. For example, based on the human migration data, we estimate that the outbreak in Beijing or Guangzhou would result in the number of infected nearly doubled.

"The COVID-19 pandemic is far more than a health crisis: it is affecting societies and economies at their core"[15]. The implemented severe control measures had not only significantly changed the course of COVID-19 spread, but also triggered substantial changes in human mobility and forced reevaluation of economic development[15–21]. We observe a long-lasting reduction of mass migration, where human movements were reduced substantially during the national lockdown and began to revive after the national reopening. The human mobility network had experienced striking structural changes as well, with the average shortest path length increasing drastically while the average degree decreasing substantially during the national lockdown. As human mobility network provides the primary pathway along which diseases are transmitted from one place to another, these significant mobility changes would contribute a lot to the prevention and control of COVID-19[21–23]. With the gradual lifting of lockdown, we also note a steady recovery of human mobility with an unequal resumption of human outflow from developed to undeveloped areas. Based on the economic performances of more than 300 cities in the first quarter of 2020, our analysis further reveals that less developed areas were likely to bear more severe economic deprivation during the COVID-19. Intuitively, areas with more cases would suffer more severe economic recession as more control efforts would be required to contain the spread of COVID-19. However, we only find that cities in Hubei province suffered the most severe economic recession, while for cities outside of Hubei, the economic development was not significantly correlated with the severity of the local epidemic. Our study helps to understand the spatial dissemination of COVID-19 and could shed light on the reevaluation of socio-economic consequences in the post-epidemic period.

2 Results

2.1 Human mobility network

The human mobility data were collected from Baidu Migration platform[24] which is curated by Chinese search engine Baidu based on its location-based services. This platform presents relative daily human movements (depicted by the Baidu Migration Index) rather than the exact number of travelers across cities and provinces in mainland China. We collected the human flow data of 366 cities at the municipal level which cover the whole territory of mainland China (See Methods section). The national human mobility network is then constructed based on the human movements across cities.

Figure 1A illustrates the aggregated human mobility network from 1 January 2020 to 23 January 2020, with nodes representing cities and edges representing the human flow between them. Cities are placed according to their geographical coordinates, and node and label sizes are proportional to the weighted degree of each city in the constructed mobility network. Cities in Hubei province and the human migration from them are highlighted in color. The corresponding human mobility data are further presented in Figure 1B, where cities in the same province are placed together and darker colors indicate larger values of human migration. For ease of visualization, only province names are shown and provincial capital cities arrive first in each provincial block. As shown in the figure, most of the large values are condensed around the diagonal in the flow matrix, which may suggest a clustered structure of human mobility where human movements primarily circulate from one city to another in the same province.

To contain the spatial dissemination of COVID-19, Wuhan was put on lockdown on 23 January 2020 (two days before the upcoming Lunar New Year). The lockdown of Wuhan had drastically reduced the population flow from Wuhan to other areas (Figure 1E). For example, compared with last year the human migration from Wuhan dropped about 75% on the first day (25 January 2020) and 90% on the third day (27 January 2020) of the Lunar New Year (in lunar calendar). Shortly after the lockdown of Wuhan, similar control measures were also implemented in other cities of Hubei province, followed by the national lockdown and stay-at-home orders. Figure 1 C-D present two snapshots of the daily human mobility network before (16 January

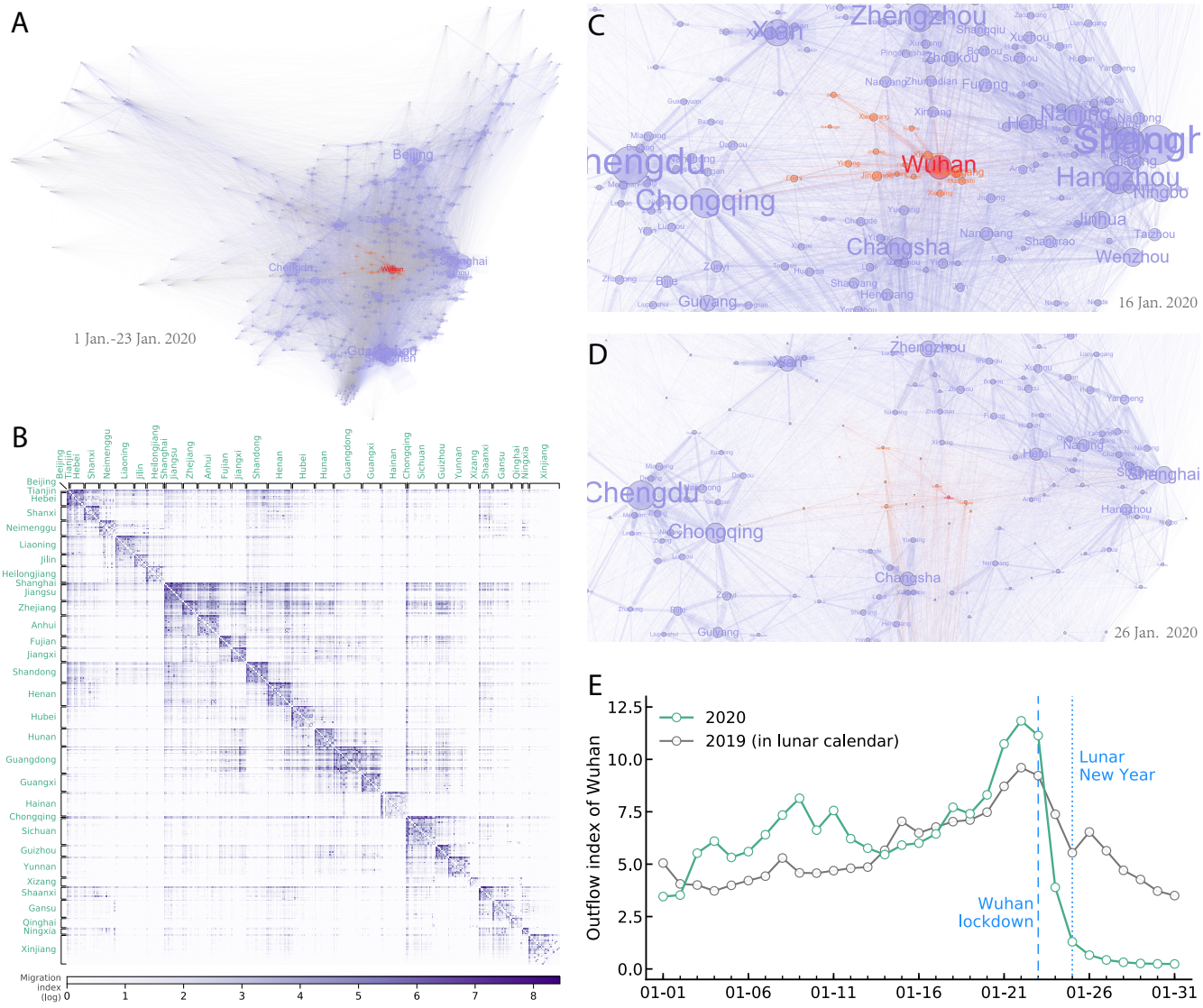

Figure 1. Human mobility network. (**A**) Human mobility network depicted by Baidu Migration Index. Node and label sizes are proportional to the weighted degree of each city in the constructed human mobility network, and edge width is proportional to the volume of human movements. (**B**) Matrix of the human mobility data corresponding to (**A**), with cities in the same province (shown in green) placed together. For ease of visualization, the raw Baidu Migration Index is multiplied by 100 and then log-transformed by $ln(x+1)$. (**C**) Snapshot of human mobility network from cities in Hubei (colored in orange with Wuhan highlighted in red) to nearby cities on 16 January 2020. (**D**) Same to (**C**) but on 26 January 2020. (**E**) Outflow index of Wuhan in January 2020 compared with that in 2019, aligned by the Lunar New Year (which is 25 January in 2020).

2020) and after (26 January 2020) the lockdown of Wuhan. Clearly, the implemented control measures had effectively cut down the social connections between Hubei and other areas.

2.2 The spatial dissemination of COVID-19

Catalyzed by the annual Spring Festival Travel Rush (which involves as many as three billion trips in a 40-day period) and the improved clinical testing capacity, the number of reported COVID-19 cases was escalating with the arrival of the Lunar New Year. As of 25 January 2020, there were only 2,010 confirmed COVID-19 cases in mainland China, but on 9 February 2020 (14 days since enforced stay-at-home orders for most provinces), 40,138 cases were confirmed: 16,902 in Wuhan, 12,729 in other cities of Hubei, and 10,507 in cities outside of Hubei province (Figure 2A). To prevent and control the local transmission of COVID-19, the Chinese New Year holiday was extended to 9 February 2020 for most provinces, which covers the estimated

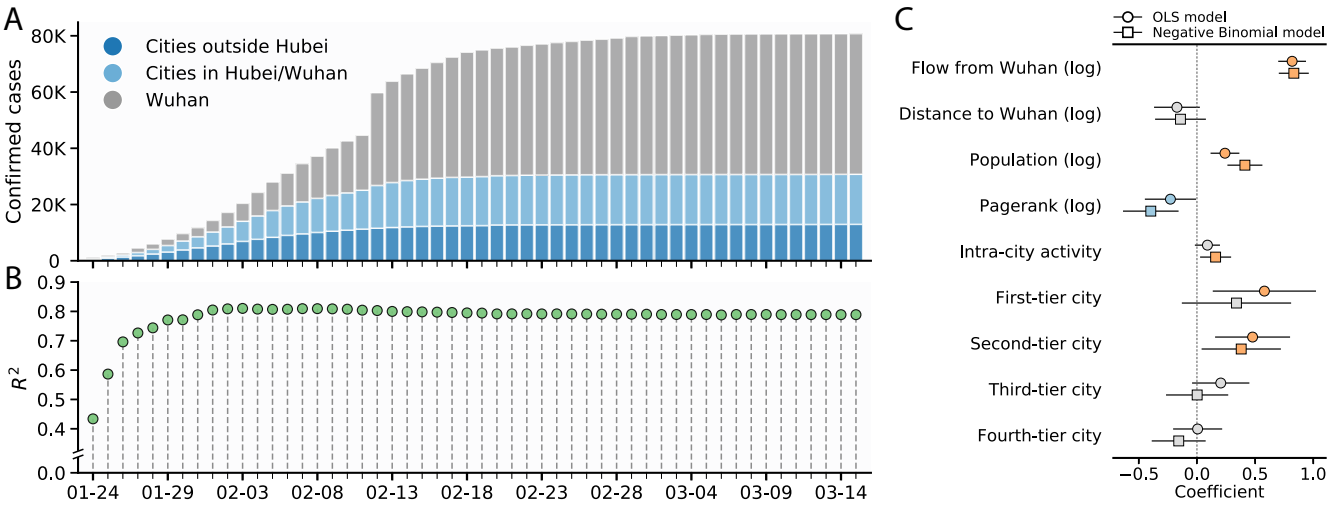

Figure 2. Cumulative COVID-19 cases. (**A**) Daily cumulative COVID-19 cases. Cities outside of Hubei province are shown in dark blue, cities inside Hubei (excluding Wuhan) are shown in light blue while the city of Wuhan is shown in grey. (**B**) Human flow from Wuhan explains the spatial distribution of COVID-19 cases. The R^2 values are obtained by OLS regressions using the number of cumulative cases (log-transformed) of each city as a function of human flow from Wuhan (log-transformed). (**C**) Coefficients from OLS regression (show in circle) and Negative Binomial regression (shown in square) are plotted, with error bars indicating 95% confidence intervals. Confidence intervals that do not cross 0 are colored.

incubation period of COVID-19. There was a surge of confirmed cases on 12 February 2020 due to the inclusion of clinical confirmation of COVID-19 (rather than merely laboratory confirmation). After March, the local transmission of COVID-19 had been greatly mitigated, and most of the cases were imported from overseas and isolated upon arrival at the border.

Consistent with previous studies[10–12], we also find that the spatial distribution of COVID-19 in mainland China can be well explained (measured by R^2) by the human flow from Wuhan (1-23 January 2020) (Figure 2B). For a given date, the R^2 value is obtained by Ordinary Least Squares (OLS) regression using the number of cumulative cases (log-transformed) on that day as a function of human flow from Wuhan (log-transformed). Specially, we achieve a R^2 value of approximately 0.8 since 31 January 2020. We further incorporate more city-specific factors in the analysis, including the distance to Wuhan, city population, intra-city activity intensity (provided by Baidu), and centrality of a city in the mobility network (measured by Pagerank[25,26]). Figure 2C illustrates the estimated coefficients for each predictor in predicting accumulative COVID-19 cases on 9 February 2020 (corresponding to the end of the extended Chinese New Year holiday for most provinces) using OLS and Negative Binomial regression models. In OLS regression, the dependent variable (i.e., number of accumulative cases) is log-transformed by $ln(x+1)$, while in Negative Binomial regression, the raw value of the dependent variable is used. As shown in the figure, we find consistent evidence that both human flow from Wuhan and city population act as significant and positive predictors ($p < 0.001$) in the case prediction. In other words, cities with larger volumes of human migration from Wuhan and more population will get more infections.

More importantly, classic complex network spreading theory would hypothesize that cities located at central positions in the mobility network are generally vulnerable to the infectious disease. However, although network centrality (measured by Pagerank) is positively associated with the number of confirmed COVID-19 cases (Spearman's $r = 0.6698$, $p < 0.001$), once the human flow from Wuhan and city population are controlled in the regression, the positive role of network centrality in the prediction of COVID-19 cases disappears (Figure 2C). This also implies the effectiveness of the implemented disease control measures across the country where most of the infected people were quarantined and isolated during the national lockdown, thereby preventing further transmission to other areas. In other words, without effective control measures, cities located at central positions would have got much more people infected due to the migration of infected people.

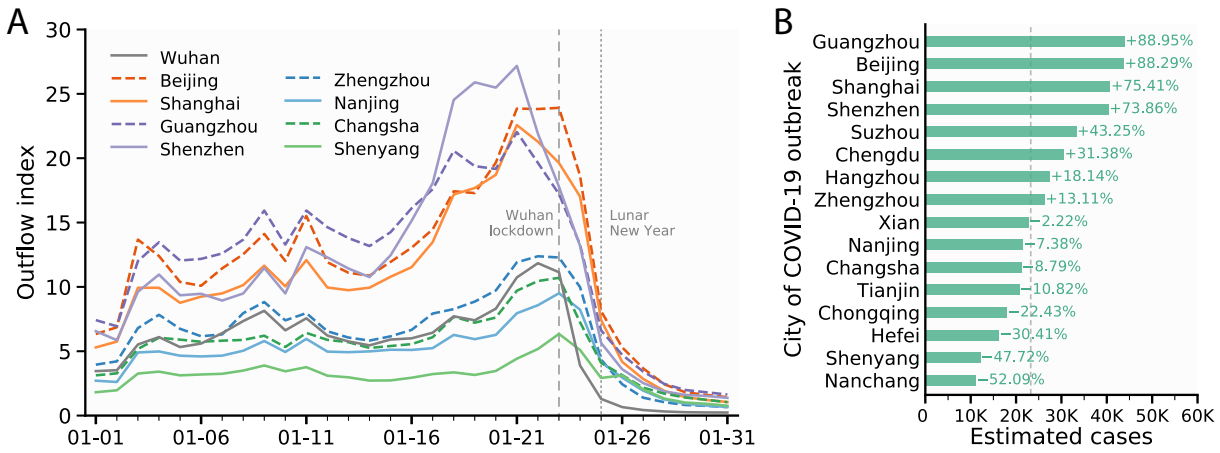

Figure 3. COVID-19 outbreak in other areas. (**A**) Outflow index of nine example cities. (**B**) Estimated accumulative COVID-19 cases if the outbreak happens in other cities. The vertical dashed line indicates the actual number of confirmed cases in cities other than Wuhan on 9 February 2020 (serves as baseline), and the number after each bar indicates the relative change of confirmed cases compared with the baseline.

2.3 COVID-19 outbreak in other cities

As we have demonstrated above, the human flow from Wuhan has largely driven the spatial distribution of COVID-19 across mainland China, therefore there could be very different consequences if the COVID-19 outbreak happens in cities other than Wuhan. Figure 3A presents the outflow index of nine example cities in January 2020, where some cities, such as Beijing, Shanghai, and Guangzhou, underwent much larger volumes of population outflow than Wuhan.

We focus on several key factors that would help the prediction of COVID-19 prevalence, including the human flow from and the distance to the outbreak city (log-transformed), city population (log-transformed), and intra-city activity intensity. For the outbreak in Wuhan (which is the real case), we fit a negative binomial regression model with the number of COVID-19 cases on 9 February 2020 set as the dependent variable and the above key factors set as the independent variables. After that, we obtain the spatial dissemination pattern of COVID-19 depicted by these key factors. Suppose that the control measures and the spatial dissemination pattern of COVID-19 remain the same. Based on the fitted model, the spatial prevalence of COVID-19 is then estimated when the outbreak city changes.

Figure 3B illustrates the estimated cumulative cases (excluding the outbreak city) as of 9 February 2020, varying with the outbreak city. The vertical dashed line indicates the actual number of confirmed cases in cities other than Wuhan on 9 February 2020 (which is 23,236) and serves as the baseline. Compared with the outbreak in Wuhan (the baseline), the relative change of cumulative cases is shown in percentage in the figure. We emphasize that our model doesn't aim to address the actual spreading process of COVID-19, but rather intends to show the relative prevalence of COVID-19 when the outbreak city changes. As shown in Figure 3B, if the COVID-19 outbreak happens in megacities like Beijing and Guangzhou, the number of confirmed cases would have been nearly doubled, but if the outbreak emerges in cities with relatively less population outflows such as Shenyang and Nanchang, we would see a shrink of infected cases by nearly half. This also suggests that where does the outbreak happen would matter a lot in the spatial dissemination of COVID-19.

2.4 Mobility changes

After the implementation of a series of control measures during the national lockdown, human mobility had undergone striking changes. Usually, we would expect a recovery of human movements since the second day of the Lunar New Year. However, due to the outbreak of COVID-19, the national migration had witnessed drastic and long-lasting shrinkage during the national lockdown (Figure 4A). For example, on the sixth day of the Lunar New Year (30 January 2020), the national migration scale dropped by nearly three quarters compared with last year. Specifically, instead of a travel surge immediately after the Lunar New Year, we observe that the national migration scale gradually decreased until the Lantern Festival (close to the end of the

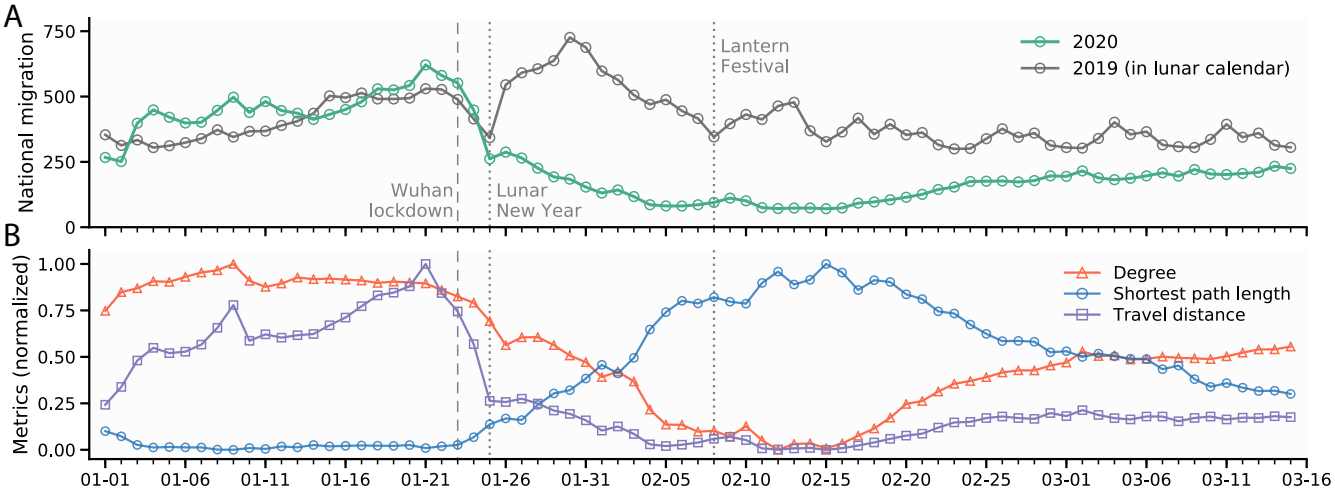

Figure 4. Human mobility during the COVID-19. (**A**) Daily national migration. (**B**) Mobility network change in terms of average degree, average shortest path length, and average travel distance. For ease of presentation, these three metrics are normalized to the range [0, 1].

extended Chinese New Year holiday). At the same time, the average geographical travel distance was also reduced substantially (Figure 4B). After the extended holiday, economic reopening was put in force orderly, and the national migration as well as long-distance travel steadily revived afterwards.

We also observe remarkable structural changes of the mobility network (Figure 4B). During the national lockdown, not only did the national migration undergo substantial reduction, but also the mobility network became less connected (measured by the average degree). Moreover, the average shortest path length of the mobility network also experienced substantial increases during the lockdown, which largely reduced the reachability of the mobility network and is able to delay the spread of virus from one city to another. Taken together, these mobility changes during the national lockdown have meaningful implications for the mitigation of disease[21,27]. After the national reopening, especially after 15 February 2020, we observe a steady recovery of the network connectivity and reachability, which indicates the gradual lifting of travel restrictions across the country.

Compared with the human migration in 2019, Figure 5 illustrates the relative inflow and outflow changes in 2020 by cities in different deciles of GDP per capita. During the national lockdown, all the cities went through similar human flow

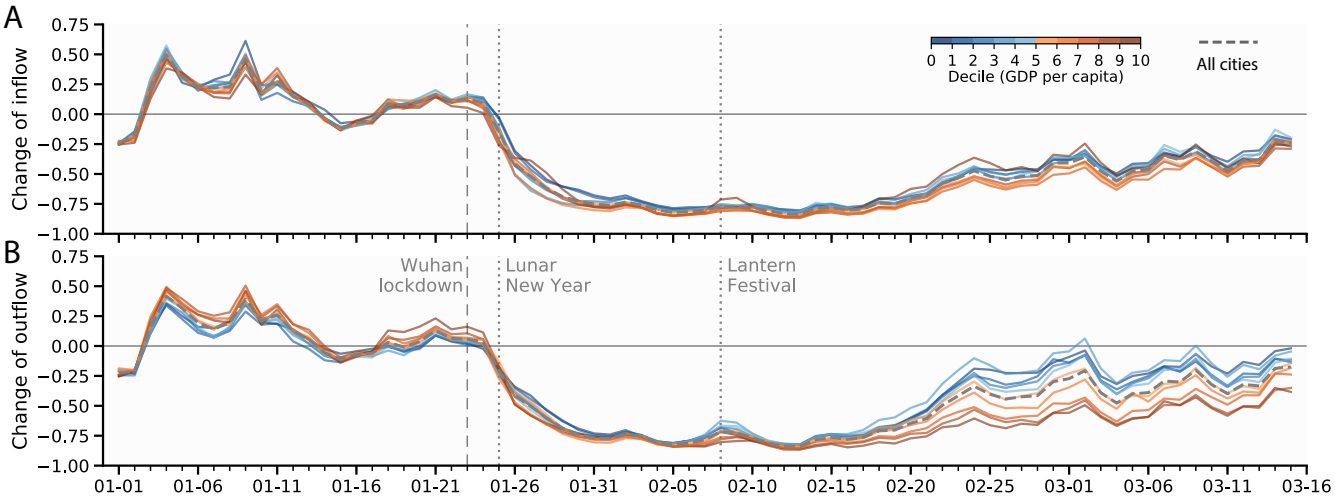

Figure 5. Relative human flow change. (**A**) Inflow and (**B**) outflow changes by cities across deciles of GDP per capita. Colors indicate cities in 1-10 deciles in terms of GDP per capita. The grey dashed curve in each panel indicates the average changes for all 366 cities.

Table 1. OLS regression results on economic development. The dependent variable is GDP growth rate in 2020Q1.

	Coef.	Std. err.	t value	p value	[95% CI]	
GDP per capita (log)	0.040	0.008	4.966	0.000	0.024	0.055
GDP growth rate (2019)	0.009	0.002	4.285	0.000	0.005	0.013
Cumulative cases (log)	0.001	0.003	0.276	0.782	-0.005	0.007
City in Hubei	-0.251	0.023	-10.689	0.000	-0.297	-0.205
Intra-city activity (aggregated)	0.0003	0.0001	3.093	0.002	0.000	0.001
Intra-city activity reduction	-0.0003	0.0001	-2.292	0.023	-0.001	-4.22e-05

observations: 330; R^2: 0.467; adjusted R^2: 0.457

reductions. After reopening, cities across deciles of GDP per capita still experienced very close inflow changes (Figure 5A), but the disparities of outflow changes were pronounced as developed areas generally suffered more severe outflow reductions than less developed areas (Figure 5B). In addition, the weekly human flow rhythm (indicated by the fluctuation of changes) emerged for both inflow and outflow after reopening, which may suggest a steady shift of human migration in the face of COVID-19.

2.5 Economic development

The prevalence of COVID-19 and the enforced control measures worldwide are plunging the world economy into the worst recession since the Great Depression[15–17]. We compile a dataset comprising the GDP records of more than 300 cities in 2019 and the first quarter of 2020 (2020Q1). To denote the economic growth of each city amid COVID-19, we divide their GDP in 2020Q1 by a quarter of their GDP in 2019. Therefore, a lower economic growth rate would indicate a more severe economic recession.

We find that less developed areas suffered more severe economic recession as cities with lower GDP per capita were likely to undergo lower rates of economic growth (Spearman's $r = 0.1185$, $p = 0.031$). This pattern becomes even clear after excluding cities in Hubei province (Spearman's $r = 0.1549$, $p = 0.006$). This finding is further validated by regression analysis where the economic growth rate is regressed on GDP per capita and several other covariates (e.g., the GDP growth rate in last year and aggregate intra-city activity intensity). Table 1 summarises the OLS regression results, where cities with higher GDP growth rates in last year would still tend to have better economic performances in the face of COVID-19 ($\beta = 0.0086$, $p < 0.001$). More importantly, after controlling several important covariates, economic growth is still significantly and positively correlated with GDP per capita ($\beta = 0.0396$, $p < 0.001$), which provides compelling evidence that less developed areas generally suffered more severe economic recession during the COVID-19. This may also imply that the economy in developed areas is more resilient to COVID-19 disruption than that of undeveloped areas.

Intuitively, cities with more infections would suffer more economic damages due to the necessary protection measures against the spread of COVID-19. However, our analysis reveals that it's city in Hubei province ($\beta = -0.2509$, $p < 0.001$) rather than the number of cumulative cases ($\beta = 0.0008$, $p = 0.782$) that is negatively correlated with economic growth. This may be explained by the fact that cities in Hubei endured the most strict lockdown measures which severely crippled their economic activities, while for most of the cities outside of Hubei, undifferentiated or similar nationwide control measures were implemented, regardless of the local epidemic incidence. Therefore, for cities outside of Hubei, more confirmed cases would probably not imply a more severe economic recession. As expected, we also find that cities with higher levels of intra-city activity ($\beta = 0.0003$, $p < 0.01$; aggregated from 1 January 2020 to 15 March 2020) and less intra-city activity reduction ($\beta = -0.0003$, $p < 0.05$; compared with last year in the lunar calendar) would have better economic performances under the threat of COVID-19.

3 Discussion

The COVID-19 pandemic is crippling our society and economy with historical levels of human movement restriction and economic deprivation. In this paper, our analysis shows that the spatial dissemination of COVID-19 in mainland China can be well explained by the human flow from Wuhan, and implies the effectiveness of the enforced lockdown across the country. Based on the spatial trajectories of human movements, it's estimated that there will be much different consequences if the COVID-19 outbreak emerged in other cities. For example, the outbreak in cities like Beijing, Shanghai, and Guangzhou would result in a much more serious epidemic spreading than the current situation, which has important implications for future epidemic prevention. We also note a remarkable reduction of human movements during the national lockdown, with significant changes of the human mobility network toward containing the spread of COVID-19.

In addition, the prevalence of COVID-19 may have unevenly affected developed and undeveloped areas as cities with lower GDP per capita and slower economic growth rate (in last year) were likely to suffer more severe economic recession under the threat of COVID-19. This finding suggests that the income and economic inequality would be enlarged without targeted interventions. Unexpectedly, for cities outside of Hubei province, more infections perhaps not suggest more severe economic deprivation. This may be induced by the fact that similar epidemic prevention and control measures were implemented for cities outside of Hubei, thus calling for differentiated and tailored epidemic control strategies based on the local epidemic.

In summary, our work contributes to a further understanding of how human mobility can be utilized to address the spatial dissemination of COVID-19 and paves a way for the application of data analytics in preventing and containing an epidemic. Our work also highlights the need to evaluate the impact of COVID-19 on our society and economy and provides practical implications for policy interventions.

Methods

Data

The human mobility data were sourced from the Baidu Migration platform[24] based on Baidu's location-based services. As the dominant search engine in China, Baidu has nearly 189 million daily active users and responses to more than 120 billion daily location service requests. Similar to recent studies[3,11], the mobility data don't indicate the absolute number of recorded trips but reflect the relative movements of people using Baidu's location-based services. We collected daily inter- and intra-city mobility data across 366 cities from 1 January to 15 March in 2020 and the corresponding period in 2019 (aligned by the Lunar New Year). For inter-city activity in 2019, only aggregated inflow and outflow data were provided for each city.

The COVID-19 data were obtained from the daily case report released by the Health Commission of each province and NetEase News[28], a professional media platform that provides timely updates and serves as a supplementary source in our study. The population and economic records in 2019 were obtained from the National Economic and Social Development Statistical Bulletin 2019 of each city. The GDP data in 2020Q1 were obtained from official reports and media coverage. For nearly 30 cities, their economic data in 2019 or 2020Q1 were not available.

Network analysis

We adopt Pagerank[25,26], a classic global network centrality measure, to quantify how important a city is in the mobility network. The human flow volume between two cities is used as the weight in the calculation of Pagerank. The average degree measures the average number of incoming and outgoing links of nodes in the mobility network. In the context of human migration, two cities are said to be close to each other if they share a large volume of human flow. As such, we use the inverse of the human flow volume to denote the "distance" of two cities along each edge in the mobility network, based on which the shortest path length from one city to another is calculated. The average shortest path length of the mobility network is thus obtained by averaging the shortest path lengths of all pairs of nodes. In practice, these network metrics were obtained using Python package `networkx`.

Statistical analysis

Most of the data processing was done by Python package `pandas` and R package `dplyr`. Spearman rank correlation was performed by Python package `scipy`; OLS regression analysis was performed by Python package `statsmodel` and R function `lm`; Negative binomial regression was performed by R package `MASS`.

Data availability

The relevant data are collected from public sources and available from the corresponding authors upon reasonable request.

References

1. Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New Engl. J. Medicine* **382**, 1199–1207 (2020).
2. Chen, S., Yang, J., Yang, W., Wang, C. & Bärnighausen, T. COVID-19 control in China during mass population movements at new year. *The Lancet* **395**, 764–766 (2020).
3. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585**, 410–413 (2020).
4. Pan, A. *et al.* Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* **323**, 1915–1923 (2020).
5. Kupferschmidt, K. & Cohen, J. Can China's COVID-19 strategy work elsewhere? *Science* **367**, 1061–1062 (2020).
6. Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
7. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489 (2009).
8. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012).
9. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
10. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–395 (2020).
11. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
12. Mu, X., Yeh, A. G.-O. & Zhang, X. The interplay of spatial spread of COVID-19 and human mobility in the urban system of China during the Chinese New Year. *Environ. Plan. B: Urban Anal. City Sci.* 2399808320954211 (2020).
13. Badr, H. S. *et al.* Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infect. Dis.* **20**, 1247–1254 (2020).
14. Kissler, S. M. *et al.* Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nat. Commun.* **11**, 4674 (2020).
15. United Nations. A UN framework for the immediate socio-economic response to COVID-19 (April 2020). https://unsdg.un.org/resources/un-framework-immediate-socio-economic-response-covid-19.
16. United Nations. Putting the UN framework for socio-economic response to COVID-19 into action (June 2020). https://www.undp.org/content/undp/en/home/coronavirus/socio-economic-impact-of-covid-19.html.
17. International Monetary Fund. World economic outlook (April 2020). https://www.imf.org/en/Publications/WEO/Issues/2020/04/14/weo-april-2020.
18. Bonaccorsi, G. *et al.* Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl. Acad. Sci.* **117**, 15530–15535 (2020).

19. Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Int. J. Surg.* **78**, 185–193 (2020).

20. Gibbs, H. *et al.* Changing travel patterns in china during the early stages of the COVID-19 pandemic. *Nat. Commun.* **11**, 5012 (2020).

21. Schlosser, F. *et al.* COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc. Natl. Acad. Sci.* **117**, 32883–32890 (2020).

22. Buckee, C. O. *et al.* Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).

23. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).

24. Baidu Migration [in Chinese]. https://qianxi.baidu.com/2020/. Accessed: June 2020.

25. Page, L., Brin, S., Motwani, R. & Winograd, T. The PageRank citation ranking: Bringing order to the web. Tech. Rep., Stanford InfoLab (1999).

26. Langville, A. N. & Meyer, C. D. A survey of eigenvector methods for web information retrieval. *SIAM Rev.* **47**, 135–161 (2005).

27. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).

28. NetEase News [in Chinese]. https://wp.m.163.com/163/page/news/virus_report/index.html. Accessed: August 2020.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61773255 and 61873167), Hong Kong RGC (GRF 11505119) and City University of Hong Kong (CCR 9360120 and HKIDS 9360163).

Author contributions

Y.Z., L.W., J.J.H.Z., and X.W. conceived the present study. Y.Z. collected and processed the data. Y.Z., J.J.H.Z., and X.W. analyzed the data. Y.Z. wrote the paper with input from L.W., J.J.H.Z., and X.W.

Competing interests

The authors declare no competing interests.

SUPPLEMENTARY INFORMATION:

The spatial dissemination of COVID-19 and associated socio-economic consequences

Yafei Zhang[1,2], Lin Wang[1], Jonathan J. H. Zhu[2,*], Xiaofan Wang[1,3,*]

1. Shanghai Jiao Tong University; 2. City University of Hong Kong; 3. Shanghai University

Note 1 - Data

1 Human mobility data

The city-level human mobility data were collected from Baidu Migration platform[1] based on Baidu's location-based services. For each city, this platform presents the top 100 outflow destinations or inflow origins in percentage (if there are less than 100 recorded destinations or origins, all the records will be shown). Based on the daily outflow or inflow migration index shown on the platform, the daily human flow data from one city to another are thus obtained.

For example, on 16 January 2020, the outflow migration index of Wuhan is 6.0027804, and 1.28% of the movements are directed to Beijing, Therefore, according to the outflow data of Wuhan, the human flow volume from Wuhan to Beijing is $6.0027804 \times 1.28\% = 7.683558912\%$. Similarly, based on the inflow data we may obtain another human flow matrix complementing the one obtained by outflow data. For example, on 16 January 2020, the inflow migration index of Beijing is 7.683179184, and 0.89% of the movements are from Wuhan. Therefore, according to the inflow data of Beijing, the human flow volume from Wuhan to Beijing is $7.683179184 \times 0.89\% = 7.683179184\%$. These two values are then averaged to depict the human flow from Wuhan to Beijing on 16 January 2020. However, if only one human flow record from city A to city B is found in the data (e.g., city B is not ranked in the top 100 destinations of city A and thus not shown in the outflow data of city A), the human flow volume from city A to city B is thus obtained based on the found record. If no records are found from city A to city B in the data, the human flow volume from city A to city B is thus denoted as 0. Based on the obtained human flow data across cities, the human mobility network is thus constructed. For ease of presentation, the raw Baidu Migration Index is multiplied by 100 in the analysis.

Figure 1A presents the human mobility network based on the aggregate population flow data from 1 January 2020 to 23 January 2020. Specifically, this network has 366 nodes and 97,650 edges. Table S1 shows the basic descriptions of the constructed human mobility network. The basic description of aggregate human flow data from Wuhan to other 365 cities is shown in Table S2. The intra-city activity, which measures the mean intra-city activity intensity (provided by Baidu migration) between 1-23 January 2020, is also shown in Table S2.

Table S1. Human mobility network description.

	count	mean	std	min	25%	50%	75%	max
Degree	366.0	533.607	101.360	211.000	461.250	514.500	594.750	730.000
Indegree	366.0	266.803	54.644	85.000	226.250	259.500	304.750	365.000
Outdegree	366.0	266.803	51.014	116.000	234.000	252.000	299.750	365.000
Weighted degree	366.0	5556.440	7211.316	21.767	1656.513	3537.837	6394.397	52915.163
Weighted indegree	366.0	2778.220	2803.648	9.974	899.161	2094.317	3695.821	20647.785
Weighted outdegree	366.0	2778.220	4608.780	11.041	750.650	1485.201	2478.662	33847.200

[1] https://qianxi.baidu.com/2020/ [in Chinese]

Table S2. Data description.

	count	mean	std	min	25%	50%	75%	max
Accumulative cases (2020-02-09)	366.0	109.667	907.214	0.000	4.000	12.000	34.750	16902.000
Accumulative cases (2020-03-01)	366.0	218.470	2590.573	0.000	5.000	15.000	44.000	49315.000
Accumulative cases (2020-04-01)	366.0	221.975	2626.571	0.000	5.000	15.000	44.000	50006.000
Human flow from Wuhan	365.0	43.162	180.383	0.006	2.116	6.702	17.232	2075.494
Distance to Wuhan (km)	365.0	1087.670	727.494	52.384	577.254	899.410	1361.106	3602.513
Population (in 10,000)	365.0	386.389	349.426	6.427	154.900	307.700	505.700	3124.320
Pagerank (weighted)	366.0	0.0027	0.0024	0.0004	0.0014	0.0021	0.0032	0.0217
Intra-city activity (mean)	366.0	5.089	0.760	1.622	4.740	5.257	5.586	6.553
GDP per capita (in 10,000 RMB)	364.0	6.041	3.460	1.416	3.563	4.811	7.693	21.043
GDP growth rate (2019)	353.0	6.397	1.971	-3.600	5.300	6.700	7.800	11.800
GDP growth rate (2020Q1)	332.0	0.826	0.092	0.389	0.802	0.843	0.875	1.035

2 Case data

The COVID-19 case data were obtained from the daily case report released by the Health Commission of each province and NetEase News[2], which serves as a supplementary source. Figure S1 presents the number of daily reported cases. There was a surge of confirmed cases on February 12 due to the change of the criteria of COVID-19 confirmation. For example, instead of laboratory confirmation, people with related COVID-19 clinical symptoms are also considered infected. As shown in the figure, most of the cases were identified before March 2020. The basic description of accumulative case data on 9 February 2020 and 1 April 2020 are also shown in Table S2.

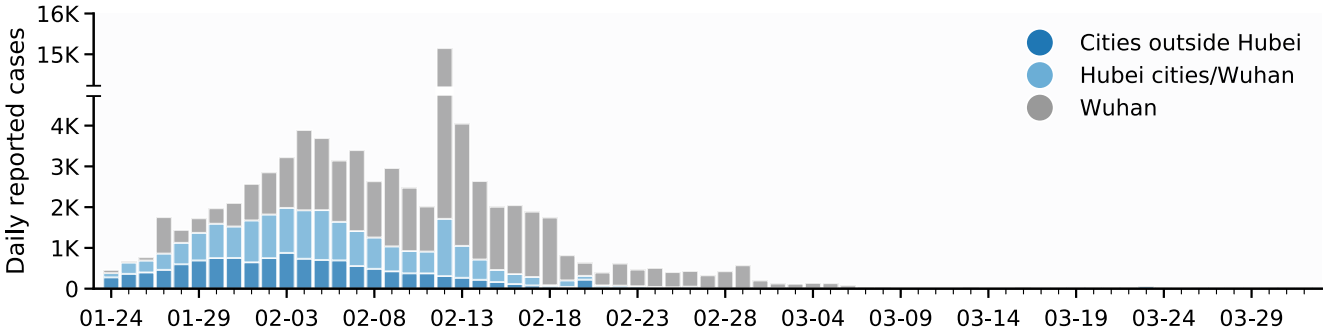

Figure S1. Number of daily reported cases.

3 Other kinds of data

Table S2 presents the basic descriptions of some key variables. Due to the constraint of data availability, there are missing values on specific data. The socio-economic data were obtained from the National Economic and Social Development Statistical Bulletin 2019 of each city and media coverage. Specifically, the GDP growth rate in 2019 indicates the relative GDP growth of each city and was obtained from the National Economic and Social Development Statistical Bulletin 2019, while the GDP growth rate in 2020Q1 was obtained by dividing the GDP in 2020Q1 by a quarter of the GDP in 2019 of each city.

In addition, Pagerank measures the global centrality of a city in the constructed mobility network using the human flow volume between cities as the weight. City tier is a synthetic evaluation of cities and was obtained from a specialized financial media platform[3]. For example, Beijing, Shanghai, Guangzhou, and Shenzhen are the four cities in Tier 1, while Wuhan, Suzhou, and other 13 cities are listed as cities in new Tier 1. Tier 1 and new Tier 1 are combined as one category (i.e., Tier 1) in this study.

[2] https://wp.m.163.com/163/page/news/virus_report/index.html [in Chinese]
[3] https://www.yicai.com/news/100648666.html [in Chinese]

Note 2 - Analysis on the spatial distribution of COVID-19

1 Regression analysis

The spatial distribution of COVID-19 in mainland China can be well explained by the human flow from Wuhan (Figure 2B). For most areas, the Chinese New Year holiday was extended to 9 February 2020, before which the national stay-at-home was implemented and after which the national reopening was put in force orderly. In this regard, the number of accumulative cases on 9 February 2020 could be an important indicator to investigate the spatial distribution of COVID-19.

After incorporating more city-specific factors, such as population and distance to Wuhan, Figure 2C in the main text presents further analysis results using OLS and Negative Binomial regression models. The pairwise Spearman rank correlation coefficients between variables can be found in Table S3. In both models, the number of accumulative cases is set as the dependent variable, except that the dependent variable is log-transformed in OLS regression due to its highly skewed distribution. The detailed regression tables (corresponding to Figure 2C) can be found in Tables S4 and S5. As shown in the tables, human flow from Wuhan and city population consistently serve as positive and significant predictors ($p < 0.001$) in predicting the number of accumulative cases in both models. Figures S2 further illustrates the number of predicted cases versus the number of confirmed cases using Negative Binomial model, where the red dots represent cities in Hubei province.

More importantly, once the human flow from Wuhan and city population are considered in the regression, the centrality of a city (measured by Pagerank in current study) no longer plays a positive and significant role in the prediction of COVID-19

Table S3. Pairwise Spearman rank correlation coefficients.

	Cumulative cases	Human flow from Wuhan	Distance to Wuhan	Population	Pagerank (weighted)
Cumulative cases	1.0***				
Human flow from Wuhan	0.881***	1.0***			
Distance to Wuhan	-0.698***	-0.783***	1.0***		
Population	0.709***	0.761***	-0.508***	1.0***	
Pagerank (weighted)	0.67***	0.77***	-0.431***	0.827***	1.0***
Intra-city activity	0.308***	0.342***	-0.357***	0.315***	0.327***

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

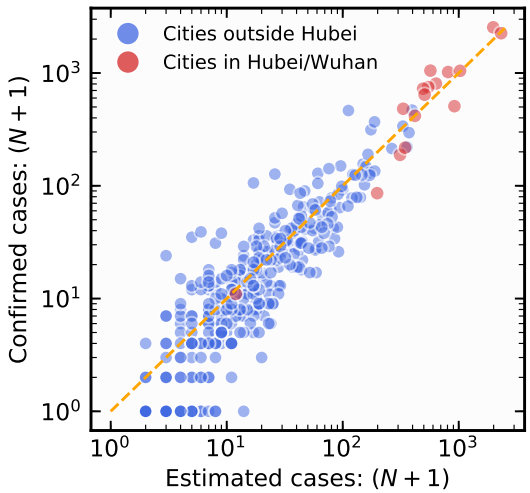

Figure S2. Number of confirmed cases versus number of estimated cases using Negative Binomial regression model. The orange line in the diagonal position indicates a perfect fit where the number of estimated cases is equal to the number of confirmed cases for each city. For ease of visualization, number of cases (N) are added by 1 (i.e., $N+1$).

Table S4. OLS regression results. The dependent variable is the number of accumulative cases on 9 February 2020, log-transformed by $ln(x+1)$.

	coef	std err	t	P> |t|	[0.025	0.975]
Human flow from Wuhan (log)	0.8187	0.061	13.440	0.000	0.699	0.938
Distance to Wuhan (log)	-0.1721	0.100	-1.716	0.087	-0.369	0.025
Population (log)	0.2410	0.063	3.811	0.000	0.117	0.365
Pagerank (log)	-0.2281	0.112	-2.042	0.042	-0.448	-0.008
Intra-city activity	0.0901	0.055	1.642	0.102	-0.018	0.198
First-tier city	0.5803	0.226	2.570	0.011	0.136	1.024
Second-tier city	0.4786	0.164	2.921	0.004	0.156	0.801
Third-tier city	0.2048	0.125	1.635	0.103	-0.042	0.451
Fourth-tier city	0.0057	0.107	0.053	0.958	-0.205	0.217
(Intercept)	-1.2712	1.457	-0.873	0.384	-4.137	1.594

Number of observations: 364; R^2: 0.832; Adjusted R^2: 0.828

Table S5. Negative Binomial regression results. The dependent variable is the number of accumulative cases on 9 February 2020.

	coef	std err	z	P> |z|	[0.025	0.975]
Human flow from Wuhan (log)	0.8333	0.066	12.673	0.000	0.704	0.962
Distance to Wuhan (log)	-0.1418	0.112	-1.271	0.204	-0.360	0.077
Population (log)	0.4125	0.077	5.369	0.000	0.262	0.563
Pagerank (log)	-0.3971	0.122	-3.257	0.001	-0.636	-0.158
Intra-city activity	0.1604	0.068	2.355	0.019	0.027	0.294
First-tier city	0.3399	0.240	1.418	0.156	-0.130	0.810
Second-tier city	0.3808	0.174	2.190	0.029	0.040	0.722
Third-tier city	0.0019	0.136	0.014	0.989	-0.266	0.269
Fourth-tier city	-0.1567	0.118	-1.323	0.186	-0.389	0.075
(Intercept)	-3.6387	1.632	-2.230	0.026	-6.837	-0.440

Number of observations: 364; pseudo R^2: 0.829

cases (Table S6; see also Table S4 and Table S5). This finding suggests that the spatial dissemination of COVID-19 in mainland China is quite different from the classic spreading process in networks where cities located at central positions would get more people infected due to the movements of infected people. This may be due to the fact that most of the infected people are quarantined and isolated during the national lockdown, thus largely preventing transmission to other areas. It also implies the effectiveness of the enforced disease control measures across mainland China as cities located at central positions would get more infections without these effective control measures.

2 Additional analysis

2.1 LASSO regression

To address the potential problem induced by the possible multicollinearity among the independent variables (e.g., Table S3) in the regression analysis, we further adopt Least Absolute Shrinkage and Selection Operator (LASSO) in the analysis. LASSO is a widely used approach in variable selection and sparse regression. It assigns L_1 penalty of the estimated coefficients (added as the regularization term) to the loss function. Specifically, LASSO will continuously shrink the coefficients of less important features to be 0 as the regularization level increases. In practice, we use the Python package `sklearn` to implement LASSO regression.

Figure S3 presents the variable trace profiles via L_1-regularized (LASSO) linear regression. As shown in the figure, human flow from Wuhan and city population are first selected before other variables and act as positive predictors, or in other words,

Table S6. Stepwise regression results. The dependent variable is the number of accumulative cases on 9 February 2020. For OLS model, the dependent variable is log-transformed.

	OLS model				Negative Binomial model			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Human flow from Wuhan (log)	1.01^{***} (0.03)			0.93^{***} (0.03)	0.99^{***} (0.03)			0.91^{***} (0.04)
Population (log)		1.01^{***} (0.06)		0.31^{***} (0.06)		0.99^{***} (0.08)		0.45^{***} (0.07)
Pagerank (log)			1.59^{***} (0.10)	-0.20^{*} (0.10)			1.72^{***} (0.13)	-0.37^{***} (0.11)
(Intercept)	0.45^{***} (0.07)	-2.94^{***} (0.34)	12.43^{***} (0.63)	-2.33^{**} (0.86)	0.68^{***} (0.08)	-1.66^{***} (0.45)	14.37^{***} (0.79)	-3.96^{***} (0.97)
R^2	0.81	0.43	0.40	0.82				
Adj. R^2	0.81	0.43	0.40	0.82				
AIC					2780.61	3310.38	3306.88	2741.73
BIC					2792.30	3322.07	3318.57	2761.22
Log Likelihood					−1387.30	−1652.19	−1650.44	−1365.87
Num. obs.	364	364	364	364	364	364	364	364

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

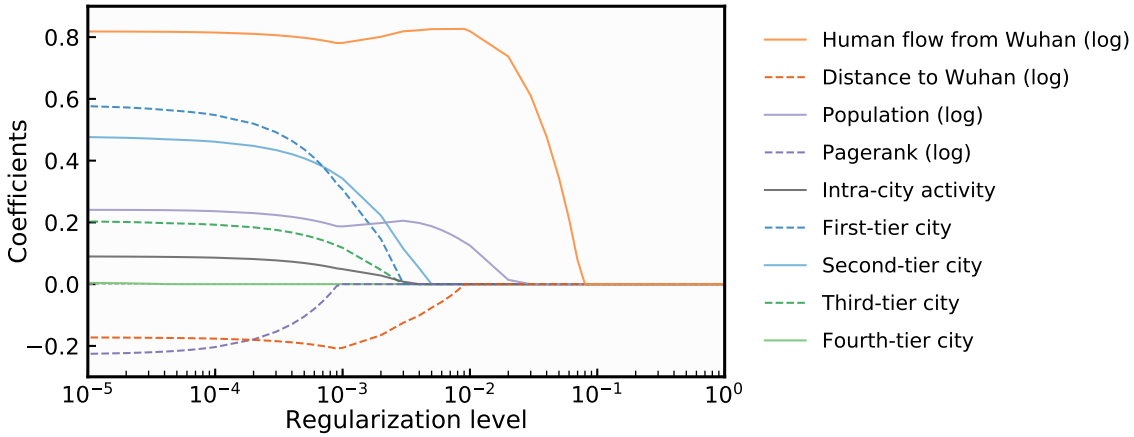

Figure S3. LASSO variable trace profiles. In the L_1-regularized linear regression, the dependent variable is the number of accumulative cases on 9 February 2020 (log-transformed). With the decrease of the regularization level, more important features are first selected.

human flow from Wuhan and city population are the two major factors in predicting the spatial distribution of COVID-19. Moreover, with the decrease of the regularization level, the centrality of the city (measured by Pagerank in the mobility network) is one of the last selected variables and consistently serves as a negative predictor in the regression. In summary, the outputs from LASSO regression provide further evidence to the findings above.

2.2 Mixed-effects model

In mainland China, most of the control measures were implemented at the province level. As such, there may exist a hierarchical structure in the data, where cities in the same province are clustered together and are non-independent with each other. To address this issue, we further adopt mixed-effects model (also called multilevel model) in the analysis. Specifically, provinces (the group factor in the data) are assigned as the random effects in the model. In practice, we adopt the R package `lme4` to implement the mixed-effects model.

Table S7 shows the regression results using linear mixed-effects model and Negative Binomial generalized linear mixed-

Table S7. Mixed-effects models. The number of accumulative COVID-19 cases on 9 February 2020 is set as the dependent variable in both models (log-transformed for linear mixed-effects model).

	Linear mixed-effects model	Negative Binomial generalized linear mixed-effects model
Human flow from Wuhan (log)	0.86^{***}	0.87^{***}
	(0.09)	(0.09)
Distance to Wuhan (log)	−0.08	−0.06
	(0.13)	(0.14)
Population (log)	0.26^{***}	0.36^{***}
	(0.08)	(0.10)
Pagerank (log)	−0.19	−0.24
	(0.13)	(0.15)
Intra-city activity	−0.05	−0.02
	(0.07)	(0.08)
First-tier city	0.35	0.16
	(0.25)	(0.28)
Second-tier city	0.35	0.26
	(0.19)	(0.20)
Third-tier city	0.14	0.03
	(0.14)	(0.14)
Fourth-tier city	0.01	−0.06
	(0.11)	(0.12)
(Intercept)	−1.12	−2.24
	(1.73)	(1.85)
AIC	748.55	2668.10
BIC	795.32	2714.86
Log Likelihood	−362.28	−1322.05
Num. obs.	364	364
Num. groups: Provinces	31	31

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

effects model. After controlling the clustered structure of the data–cities in the same province are clustered together, human flow from Wuhan and city population remain positive and significant predictors ($p < 0.001$) in the case prediction. Consistent with the main analysis, the centrality of a city (measured by Pagerank) still doesn't play a positive and significant role in the case prediction.

2.3 Analysis on other dates

To investigate the factors that help to explain the spatial dissemination of COVID-19, we present regression results when the dependent variable is the number of accumulative cases on 9 February 2020 in the main text and previous sections (Figure 2C; Tables S4, S5 and S7). In this section, we present additional regression analysis to predict the spatial distribution of COVID-19 cases on 1 March 2020 and 1 April 2020.

Table S8 shows the regression results using OLS and Negative Binomial models. As shown in the table, the primary findings in the main text remain the same. Specifically, human flow from Wuhan and city population still act as significant and positive predictors in the case prediction on 1 March 2020 and 1 April 2020. The centrality of a city (measured by Pagerank) in the mobility network doesn't play a positive and significant role in the case prediction.

Table S8. Statistical results for case prediction on 1 March 2020 and 1 April 2020. The number of accumulative COVID-19 cases on the given date is set as the dependent variable in each regression model. For OLS models, the dependent variable is log-transformed.

	OLS model		Negative Binomial model	
	Model 1 (1 March 2020)	Model 2 (1 April 2020)	Model 3 (1 March 2020)	Model 4 (1 April 2020)
Human flow from Wuhan (log)	0.86***	0.87***	0.78***	0.79***
	(0.07)	(0.07)	(0.08)	(0.08)
Distance to Wuhan (log)	−0.14	−0.13	−0.23	−0.21
	(0.11)	(0.11)	(0.13)	(0.13)
Population (log)	0.28***	0.28***	0.48***	0.49***
	(0.07)	(0.07)	(0.09)	(0.09)
Pagerank (log)	−0.25*	−0.26*	−0.34*	−0.36*
	(0.12)	(0.12)	(0.14)	(0.14)
Intra-city activity	0.09	0.09	0.02	0.02
	(0.06)	(0.06)	(0.07)	(0.07)
First-tier city	0.46	0.54*	0.08	0.19
	(0.25)	(0.25)	(0.28)	(0.28)
Second-tier city	0.40*	0.41*	0.16	0.16
	(0.18)	(0.18)	(0.20)	(0.20)
Third-tier city	0.19	0.20	−0.05	−0.05
	(0.14)	(0.14)	(0.16)	(0.16)
Fourth-tier city	−0.04	−0.04	−0.31*	−0.31*
	(0.12)	(0.12)	(0.13)	(0.14)
(Intercept)	−1.68	−1.98	−1.93	−2.22
	(1.61)	(1.61)	(1.85)	(1.85)
R^2	0.81	0.81		
Adj. R^2	0.81	0.81		
AIC			2960.11	2967.53
BIC			3002.98	3010.40
Log Likelihood			−1469.05	−1472.76
Num. obs.	364	364	364	364

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table S9. Pairwise Spearman rank correlation coefficients.

	GDP growth rate (2020Q1)	GDP per capita	GDP growth rate (2019)	Cumulative cases	City in Hubei	Intra-city activity (aggregated)
GDP growth rate (2020Q1)	1.0***					
GDP per capita	0.1185*	1.0***				
GDP growth rate (2019)	0.2849***	0.1244*	1.0***			
Cumulative cases	-0.0708	0.344***	0.1797***	1.0***		
City in Hubei	-0.3253***	0.0982	0.1725**	0.336***	1.0***	
Intra-city activity (aggregated)	0.2384***	-0.231***	0.0908	-0.0745	-0.2783***	1.0***
Intra-city activity reduction	-0.2044***	0.2809***	-0.1165*	0.36***	0.3234***	-0.5794***

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Note 3 - Economic growth

Table S9 shows the pairwise Spearman rank correlation coefficients between variables. Table S10 presents the regression results on economic growth in 2020Q1. Specifically, we use the GDP growth rate in 2020Q1 to denote the economic growth during the COVID-19. Intuitively, a lower GDP growth rate would indicate a more severe economic recession. City in Hubei denotes whether a city is located in Hubei province or not. We use the number of accumulative cases on 9 February 2020 to denote the severity of the local epidemic. The activity intensity measures the aggregate intra-city activity intensity (provided by Baidu migration) from 1 January 2020 to 15 March 2020, while the activity reduction measures the relative reduction of the intra-city activity compared with last year (2019).

As shown in Table S10, the inclusion of city in Hubei significantly improves the model fitting with R^2 value improved from 0.278 to 0.467 (ANOVA analysis, $Fstatistic = 114.26$, $p < 0.001$). More importantly, once the city in Hubei is controlled, the number of cumulative cases is no longer a significant and negative factor in the prediction of economic growth. This may suggest that cities in Hubei suffered the most severe economic recession, while for cities outside of Hubei, more confirmed cases would probably not imply more severe economic recession.

Table S10. Stepwise OLS regression on economic growth.

	Dependent variable:	
	GDP growth rate (2020Q1)	
	Model 1	Model 2
GDP per capita (log)	0.055^{***}	0.040^{***}
	(0.009)	(0.008)
GDP growth rate (2019)	0.006^{**}	0.009^{***}
	(0.002)	(0.002)
Cumulative cases (log)	-0.014^{***}	0.001
	(0.003)	(0.003)
City in Hubei		-0.251^{***}
		(0.023)
Intra-city activity (aggregated)	0.0005^{***}	0.0003^{**}
	(0.0001)	(0.0001)
Intra-city activity reduction	-0.0004^{**}	-0.0003^{*}
	(0.0002)	(0.0001)
(Intercept)	0.619^{***}	0.637^{***}
	(0.046)	(0.040)
R^2	0.278	0.467
Adjusted R^2	0.267	0.457
Observations	330	330

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$