

Network analysis of population flow among major cities and its influence on COVID-19 transmission in China

Jie Liu, Jingyu Hao, Yuyu Sun, Zhenwu Shi*

Northeast Forestry University, School of Civil Engineering, 26 Hexing Road, Harbin, China

ARTICLE INFO

Keywords:
Population flow network
COVID-19 transmission
Network analysis
Importation risk

ABSTRACT

Large-scale and diffuse population flow amplifies the localized COVID-19 outbreak into a widespread pandemic. Network analysis provides a new methodology to uncover the topology and evolution of the population flow and understand its influence on the early dynamics of COVID-19 transmission. In this paper, we simulated 42 transmission scenarios to show the distribution of the COVID-19 outbreak across China. We predicted some original (Guangzhou, Shanghai, Shenzhen) had higher total aggregate population outflows than Wuhan, indicating larger spread scopes and faster growth rates of COVID-19 outbreak. We built an importation risk model to identify some major cities (Dongguan and Foshan) with the highest total importation risk values and the highest standard deviations, indicating the core transmission chains (Dongguan-Shenzhen, Foshan-Guangzhou). We built the population flow networks to analyze their Spatio-temporal characteristics and identify the influential sub-groups and spreaders. By removing different influential spreaders, we identified Guangzhou can most influence the network's topological characteristics, and some major cities' degree centrality was significantly decreased. Our findings quantified the effectiveness of travel restrictions on delaying the epidemic growth and limiting the spread scope of COVID-19 in China, which helped better derive the geographical COVID-19 transmission related to population flow networks' structural features.

1. Introduction

Nowadays, coronavirus (COVID-19) is spreading quickly and globally. As of 08:00 am (GMT + 8) on 29 July 2020, it has affected 216 countries and territories, 16,523,815 have been confirmed positive for COVID-19, and 655,112 have died (World Health Organization, 2020). As a contact transmission, the COVID-19 transmission is mostly due to person-to-person contact (Baker et al., 2020). Therefore, for an epidemic with the human-to-human transmission characteristic, extensive population flow has substantially increased social contacts in public, which caused COVID-19 to reach essentially everywhere (Bherwani et al., 2020). Moreover, many patients with asymptomatic infection had traveled to other regions, which caused the confirmed case numbers to increase exponentially at the early stage of the COVID-19 outbreak (Tian et al., 2020). Enforcing strong travel restrictions was urgent, necessary, and effective. It might prevent further seeding of this virus to wider regions (Giordano et al., 2020; Lai et al., 2020). More than 130 countries have introduced some forms of travel restriction since the COVID-19 outbreak began, including screening, quarantine, and banning travel

from high-risk areas. Therefore, it is crucial to understand how the population flow influences COVID-19 transmission (Chen et al., 2020; Kraemer et al., 2020). Compared with other affected countries, China was the first country to respond to COVID-19. In China, COVID-19 was first identified and reported in Wuhan. On 31 December 2019, before the Chinese Lunar New Year (began on 24 January 2020), the outbreak of COVID-19 spread rapidly from Wuhan, a city of 11 million inhabitants and the most significant transport hub in Central China. Further spatial spread of COIVD-19 was of great concern because of the upcoming Spring Festival Holiday (from 24 January 2020 to 31 January 2020). This holiday is the most celebratory time of the year in China, during which a massive human migration takes place as individuals travel back to their hometowns (Chen et al., 2020). National Development and Reform Commission stated that the annual spring migration (called chunyun in China) often lasts 40 days. During 2020 spring migration (from 10 January 2020 to 18 February 2020), there had been typically 3 billion travel movement between urban and rural areas (Chinazzi et al., 2020). This massive travel movement formed the topology of the population flow network, which connected most major cities to initiate the

* Corresponding author.

E-mail address: shizhenwu@126.com (Z. Shi).

national transmission chains of COVID-19 in China (Shi et al., 2020; Wu et al., 2020).

Focus on the COVID-19 outbreak in China, many epidemiology-inspired models, including susceptible-infectious-recovered (SIR), susceptible-infectious-susceptible (SIS), and susceptible-exposed-infectious-recovered (SEIR) models, had been built to study the spatial spreading pattern of COVID-19 (He et al., 2020; Hou et al., 2020; Yang et al., 2020). However, these models were just based on very simplified assumptions of the population flow, which just gave rise to the model-dependent transmission pattern of COVID-19. They might fail to account for the actual dynamic of population flow affecting the spreading of COVID-19 in a real setting (Jia et al., 2020). As a consequence, there was a critical need for data-driven quantitative studies. Such studies used the actual Spatio-temporal evolution of population flow to precisely portray the spatial variation in COVID-19 transmission, which ultimately provided policymakers with valuable information on how best to target public health interventions to some certain affected cities or areas. It was essential to refine future epidemiological models (Wiwanitkit & Joob, 2020; Yue et al., 2020).

These facts motivated us to address and solve some central research questions in this paper. First, during our observation period (from 1 January 2020 to 23 February 2020), we set Wuhan as the baseline scenario, simulated 42 transmission scenarios by assuming different major cities as the original epicentres, which showed the potential implications of COVID-19 outbreak in different major cities of China. This could identify the specific topological properties of population flow, characterize the transmission patterns and outbreak sizes of COVID-19 across China, if the original epicentre was not Wuhan but other 42 major cities. Second, we built an importation risk model by conceptualizing the importation risk value as the combination of the aggregate population inflow and the cumulative number of confirmed COVID-19 cases, which quantitatively examined the effectiveness of the travel restrictions on COVID-19 transmission across China. Since the population flow can be naturally investigated from the network-theoretic perspective, network science produced an important advance in analyzing how the population flow network's structure affected the spreading performance of COVID-19 and identifying which cities can maximize the speed and scope of COVID-19 transmission. Third, we built the population flow networks to discover the underlying patterns of population flows among 42 major cities, which could analyze the population flow among multiply epicentres associated with the COVID-19 transmission. Moreover, we examined the network structure via density and average clustering coefficient at the system level. This could prove whether the topological structure and evolution of the population flow network influenced the COVID-19 transmission. We then identified the influential sub-groups and the influential spreaders via k-core and degree centrality at the individual node level. This could demonstrate which sub-groups or major cities form the cores in the population flow network, facilitating COVID-19 transmission efficiency. When the time was taken into account, the temporal network analysis of the population flow networks highlighted the changes of these particular properties over time, which can provide insights into understanding the dynamics of COVID-19 transmission. Our findings filled the research gap in data-driven insights into the population flow network's topological structure and how it is associated with the COVID-19 transmission.

2. Data and methods

2.1. Data sources

Our research was based on a data set that included case reports, population flow, and public health interventions. Our study areas included 43 major cities (including Wuhan) in 31 provinces and regions of China. According to Chinese cities' ranking list published by The New First-tier City Institute, we selected all First-tier and New First-tier, some Second-tier cities. They were 4 First-tier cities (Beijing, Shanghai,

Guangzhou, Shenzhen), 15 New First-tier Cities (Chengdu, Hangzhou, Chongqing, Wuhan, Xi'an, Suzhou, Tianjin, Nanjing, Changsha, Zhengzhou, Dongguan, Qingdao, Shenyang, Ningbo, Kunming), 20 Second-tier cities (Wuxi, Foshan, Hefei, Dalian, Fuzhou, Xiamen, Harbin, Jinan, Wenzhou, Zibo, Nanning, Changchun, Shijiazhuang, Tangshan, Guiyang, Nanchang, Taiyuan, Yantai, Lanzhou, Haikou). We added other 4 provincial capital cities (Xining, Hohhot, Urumqi, Yinchuan), which were not included in this ranking list of First-tier, New First-tier, and Second-tier cities. Due to lacking data in the daily confirmed COVID-19 cases or the daily population inflow and outflow indexes of some Second-tier cities, the study areas were limited to 43 major cities (including Wuhan) across China, which were the industrial, financial, and commercial hearts of China. And the gross domestic product (GDP) of these study areas made up more than 45% of the national GDP in 2019 (National Bureau of Statistics, 2019). From 1 January 2020 (World Health Organization declared COVID-19 as a public health emergency of international concern) to 23 February 2020 (The rise in incidence was halted and reversed in China), as the first 54 days of COVID-19 outbreak in China, our observation period included 40 days of spring migration, 15 days before the Chinese Lunar New Year on 25 January 2020, and 25 days afterward. As of 23 February 2020, China Center for Disease Control and Prevention announced there were 77,150 confirmed COVID-19 cases in China, of which 46,607 occurred in Wuhan, and 5,500 occurred in 42 other major cities we studied. Fig. 1 showed the geographical distribution of the confirmed COVID-19 cases in 43 major cities. When Wuhan was quarantined from 23 January 2020, the cumulative number of confirmed COVID-19 cases in these major cities was reported each day (Fig. 2). Considering the different population densities, we focused on the cumulative number of the confirmed COVID-19 cases (per 1 million population) to indicate the growth rate of infections in these major cities. Until 23 February 2020, the cumulative number of confirmed COVID-19 cases (per 1 million population) in Wuhan (4146) was more than 70 times higher than that in other top ten cities. These top ten cities, which were ranked according to the cumulative number of the confirmed COVID-19 cases (per 1 million population), included three first-tier cities (Beijing, Guangzhou, Shenzhen), two famous winter resort cities (Harbin, Haikou), and two essential transport hubs (Hefei, Changsha) in China. The epidemic growth rates of 43 major cities increased and reached a peak between 25 January 2020 and 28 January 2020, which was due to the increased testing capacity in China. There was a second peak in most major cities between 2 February 2020 and 4 February 2020. On 23 February 2020, except Wuhan, Chongqing, and Dongguan, the daily confirmed cases numbers in other 40 major cities had reduced close to zero (National Health Commission of the People's Republic of China, 2020), indicated the epidemic curves of most major cities were flattened.

For investigating the effectiveness of travel restrictions on COVID-19 transmission in China, we compared the population flow among 43 major cities (including Wuhan) during the same observation period in 2019 (12 January 2019–8 March 2019) and 2020 (1 January 2020–23 February 2020). This data was extracted from Baidu Inc. (Baidu Migration Map-Big data on Spring Migration). Fig. 3 showed the daily population inflow and outflow indexes of Wuhan and other top ten cities (4 first-tier cities, three transport hubs, and three developed coastal cities in southern and eastern China) with highest average daily population inflow and outflow indexes. In Fig. 3, the grey dotted line represented the day of Chinese Lunar New Year (6 February 2019, 25 January 2020). Because the spring migration was set according to the lunar calendar, we identified the observation periods in 2019 and 2020 according to the same lunar dates. Before the Chinese Lunar New Year, that was historically the peak period for spring migration with the increasing daily population outflow indexes of Wuhan and other top ten cities in 2019 and 2020 (Fig. 3c). However, in 2020, the average daily population outflow indexes of these cities were more than 1.1 times that in 2019, especially in Wuhan (1.2 times). After the Chinese Lunar New Year, that was also historically the peak period for traveling and

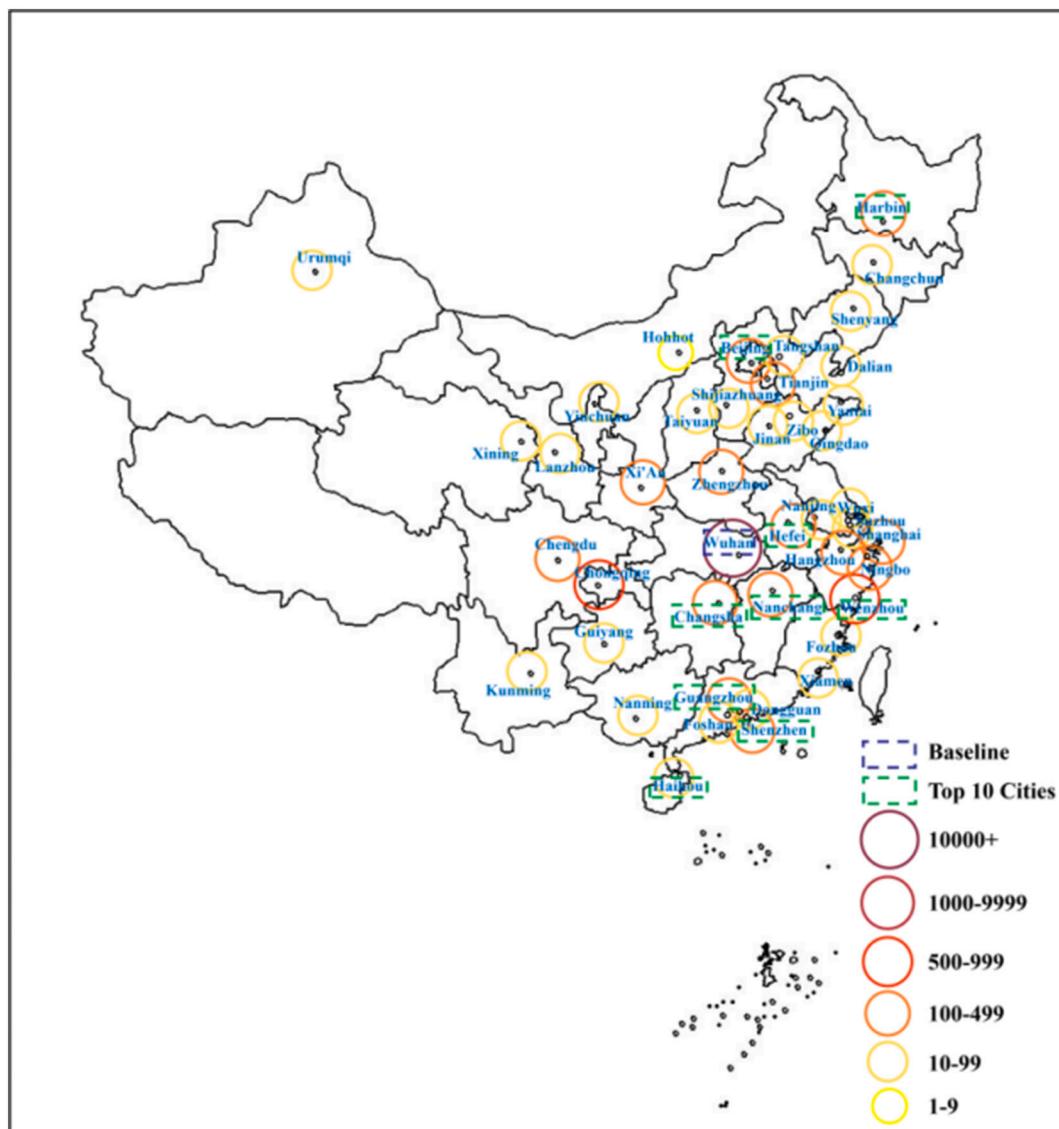


Fig. 1. Geographical distribution of the confirmed COVID-19 cases in 43 major cities until 23 February 2020.

returning to work, with the increasing daily population inflow indexes of Wuhan and other top ten cities just in 2019. Travel restrictions and large-scale control measures were implemented in China from 23 January 2020, which caused a rapidly decreasing daily population inflow and outflow indexes of these cities. During this period, the average daily population inflow and outflow indexes of these cities were 30% of that in 2019, especially in Wuhan (only 9%).

2.2. Simulating the transmission scenarios

The population outflow from Wuhan was proved to be the primary driver of COVID-19 transmission before this city was quarantined on 23 January 2020 (Yue et al., 2020). However, there is no guarantee that the second or third wave of COVID-19 will still origin from Wuhan in the future, such as a cluster of confirmed COVID-19 cases in Shulan (11 May 2020), Beijing (13 June 2020), Dalian (23 July 2020). To test the different original epicentres' contributions to seeding COVID-19 elsewhere in China, we simulated 42 transmission scenarios by assuming 42 major cities as the original epicentres, respectively. We then compared these transmission scenarios with the baseline scenario (Wuhan). During the whole observation period (1 January 2020–23 February 2020), the key date was 23 January 2020, when Wuhan was quarantined.

According to the critical date, we divided the whole observation period into two phases. Phase one (1 January 2020–23 January 2020) was when the spread of COVID-19 started from 'one seed'. Phase two (24 January 2020–23 February 2020) was when the local outbreak shifted into the national transmission. During phase one, the virus transmitted from person to person contagiously in an original epicentre and may eventually spread to other major cities due to the population outflows from the original epicentre. Moreover, there was a relative lack of awareness of this new virus and few countermeasures preventing its spread. COVID-19 should thus have spread randomly from the original epicentre to other major cities. With the imposition of the quarantine, population outflow from Wuhan almost completely stopped, manifested in a reduction of 84% in inter- and intra-provincial population flow (Kucharski et al., 2020). Therefore, we assumed that no people entered any other major cities from an original epicentre after 23 January 2020. In each transmission scenario, during phase one (1 January 2020–23 January 2020), we defined the aggregate population outflow from an original epicentre to a specific major city as the sum of population outflow indexes. Because the aggregate population outflow from an original epicentre exhibited a highly and progressively stronger correlation with infection prevalence in destination locations over that time (Tian et al., 2020), the geographical aggregate population outflow

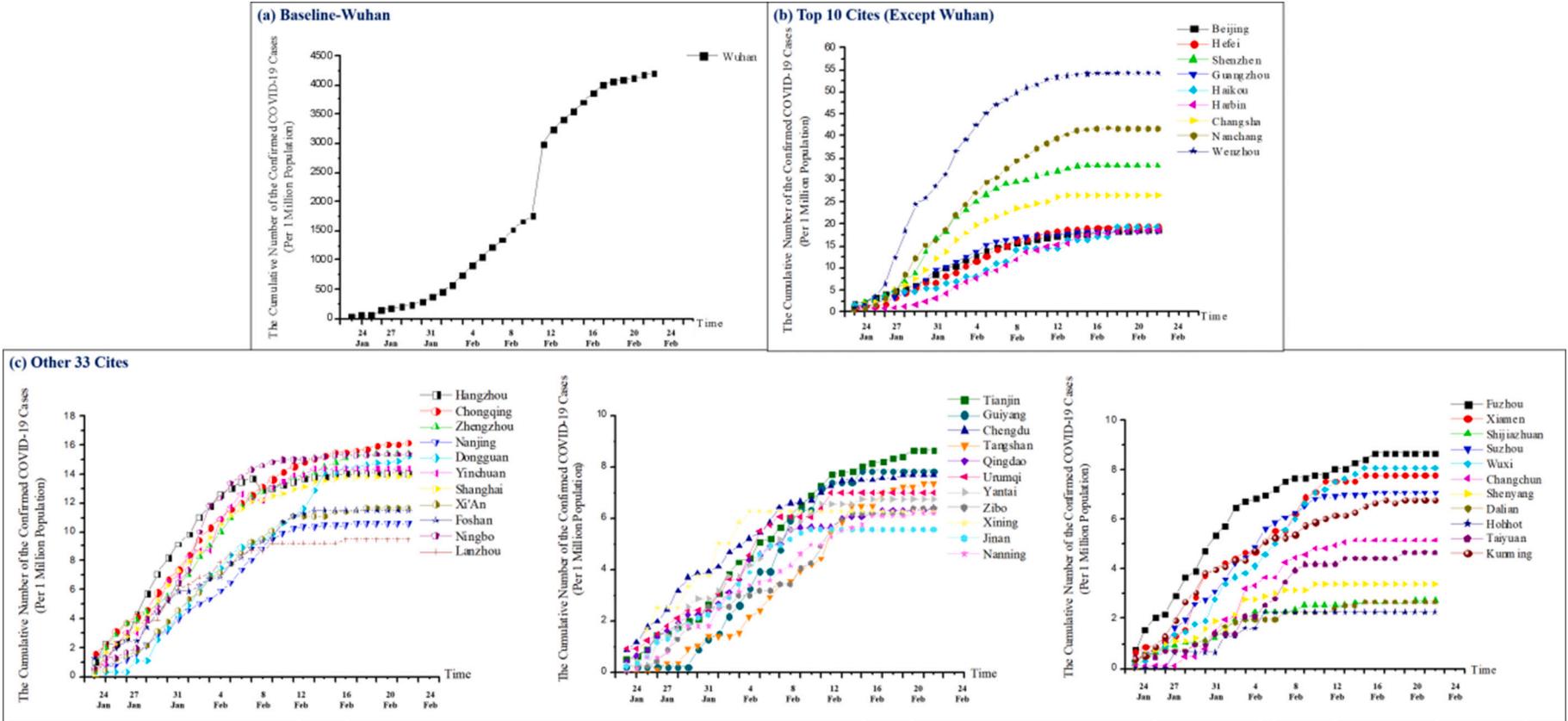


Fig. 2. The cumulative number of the confirmed COVID-19 cases (per 1 million population) in 43 major cities from 23 January 2020 to 23 February 2020.

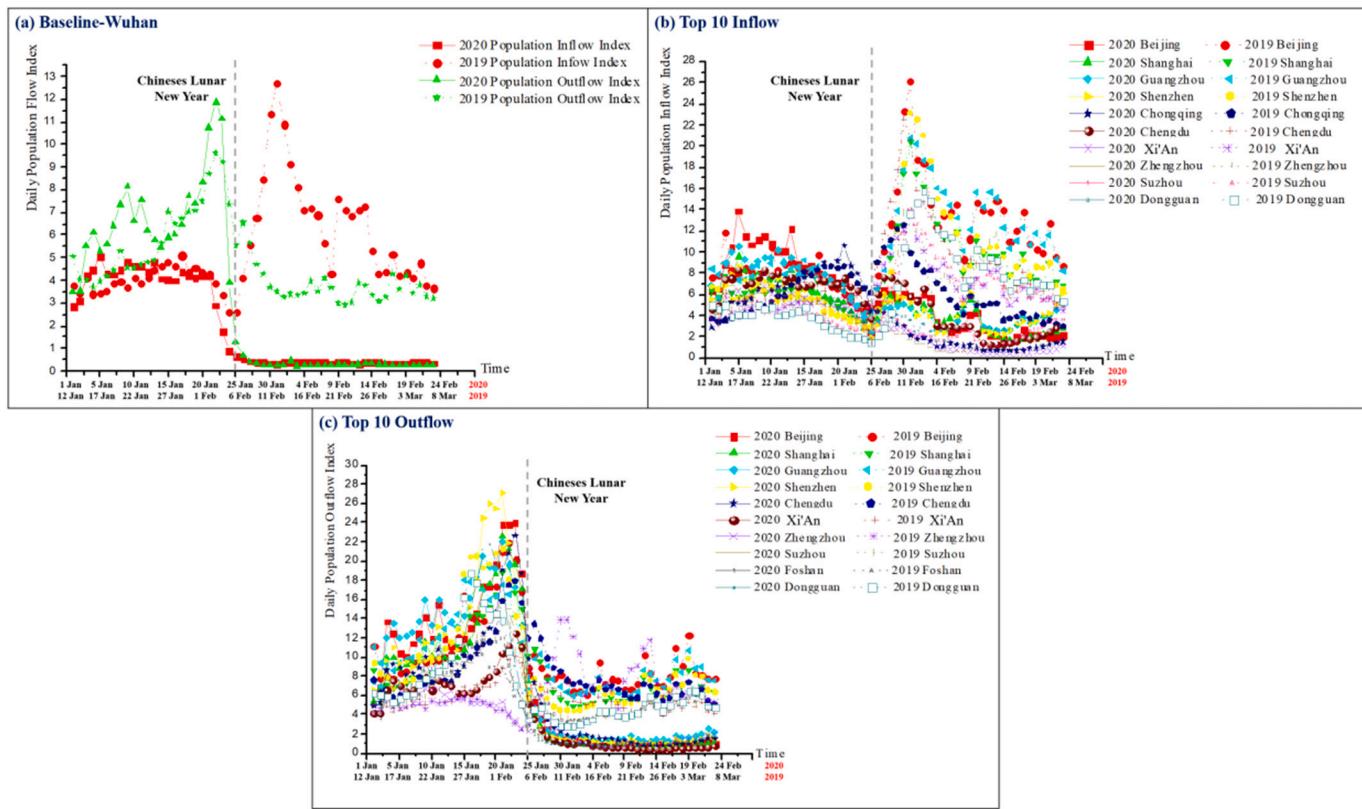


Fig. 3. The daily population inflow and outflow indexes of Wuhan and other top ten cities during the same observation period in 2019 and 2020.

anticipated the following location, intensity, and timing of the COVID-19 outbreak in other major cities across China. These simulated transmission scenarios can statistically predict the different distribution of infections with COVID-19 across China.

2.3. Building the importation risk model

After the quarantine of Wuhan and other 14 cities in Hubei Province (23 January 2020), there was little impact of the population outflows from Hubei (especially Wuhan) on the spread of COVID-19 across China (Xu, Chen, et al., 2020). Notably, the confirmed COVID-19 cases had been reported in most major cities. Therefore, there were multiple epicentres across China after 23 January 2020, not just Wuhan. The dynamics of COVID-19 transmission has shifted from the local outbreak into the national transmission. To investigate these meaningful shifts through time, we developed an importation risk model to evaluate the total importation risk values of 42 major cities during phase two (24 January 2020 to 23 February 2020) as follow:

$$\text{IR}(t|x_{ij}) = \text{PI}(t|x_i, x_j) \times \text{CC}(t|x_j) \quad (1)$$

$$\text{TIR}(t|x_i) = \sum_{j=1}^n \text{IR}(t|x_{ij}) \quad (2)$$

In which $\text{IR}(t|x_{ij})$ is the importation risk value of the city x_i from the city x_j until time t , $\text{PI}(t|x_i, x_j)$ is the hazard function describing the aggregate population inflow indexes of the city x_i from the city x_j until time t , $\text{CC}(t|x_j)$ is the vulnerability function describing the cumulative number of confirmed COVID-19 cases (per 1 million population) in the city x_j until time t , $\text{TIR}(t|x_i)$ is the total importation risk value of the city x_i until time t , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$, n is the number of the major cities.

In this model, we focused on the aggregate population inflow, rather than daily population inflow, to eliminate the incubation period's effect.

However, the COVID-19 pandemic is a dynamic process. We calculated and compared each city's daily total importation risk value to represent the effect of the dynamic population inflow on COVID-19 transmission. Besides that, we just used the aggregate population inflow, not considering the aggregate population outflow, because we assumed that the aggregate population inflow to the city x_i from the city x_j was equal to the aggregate population outflow from the city x_j to the city x_i . Therefore, a higher total importation risk value of a specific major city suggested more confirmed COVID-19 cases imported into this city, either due to the higher aggregate population inflow to this city or more cumulative confirmed COVID-19 cases in the epicentres. Our importation risk model's logic differed from classic epidemiological models that relied on the 'well-mixed' population, which assumed each individual was equally likely to encounter every other infected individual in each city. In reality, infections varied considerably in the frequency, distance, and nature of their flows in different cities (González et al., 2008; Song et al., 2010; Vazquez-Prokopec et al., 2013), in ways that related to COVID-19 transmission (Perkins et al., 2013). By statistically deriving the population inflows' temporal and spatial variation, our importation risk model can parsimoniously capture the dynamic shifts in distribution of COVID-19 across 42 major cities over time. As a validity check, we calculated the correlation between the total importation risk values and the cumulative number of confirmed COVID-19 cases (per 1 million population) of 42 major cities during phase two. Consistent with our hypothesis, all of Spearman's correlation coefficients were above 0.877, which indicated a highly correlation between these two parameters. Therefore, this importation risk model becomes more predictive of local COVID-19 outbreak, which can also act as an early warning index to initiate the local transmission chains of COVID-19.

2.4. Analyzing the population flow network and its influence on COVID-19 transmission

During phase two (24 January 2020–23 February 2020), some cities

were defined as the influential spreaders due to their high speed and the wide scope of COVID-19 transmission. Their unique transmission patterns were attributed to their dominant population flow networks (Yang & Jung, 2020). To identify which cities were the influential spreaders, and capture the dynamics of their topological properties, we modeled the population flow networks as graphs whose nodes represented the major cities and whose links represented the population inflow among these major cities. Mathematically, the graphs as the directed and weighted networks can be represented by $G = (V, E, W)$, where V was a set of N nodes, E was a set of M links, each graph contained the same number of nodes, and different links. W was the adjacency weighted matrix, and the weights of the connected links can reflect the importation risk values of nodes ($IR(t|x_{ij})$). Then we adopted a set of commonly used topological metrics to analyze the structural evolution of the population flow networks during phase two, both at the system level and the individual node level. At the system level, density and average clustering coefficient were used to assess the structural integration of the entire population flow network (Giustolisi et al., 2020). At the individual node level, degree centrality and K-core were used to address the node importance quantification and identify influential sub-groups and spreaders (Tiwari et al., 2020). All network analysis was conducted by using the UCINET and Pajek software. (1) Density was a direct and widely-used topological measure of the network. In the population flow networks, the most connected network was usually believed to be responsible for the highest importation risk. (2) Average clustering coefficient was a standard metric for quantifying the connection of the network. In the population flow networks, the highest average clustering coefficient indicated that the major cities tended to form tight clusters with a high frequency of connectivity. (3) Degree centrality provided an answer to the question ‘which nodes were central in the network?’ (Ghalmane et al., 2019). In the directed population flow networks, degree centrality (including out-degree, in-degree) was used to measure the major cities’ different capabilities to spread COVID-19 and identify which major cities as the influential spreaders may be cohesive in the widespread of COVID-19. (4) K-core was defined as a hierarchical set of nodes based on the number of links and the degree of connections (Burleson-Lesser et al., 2020; Liu et al., 2015). In the population flow networks, k-core was used to identify the connected sub-structures and the overall relations of COVID-19 transmission within the population flow network. Degree centrality alone was not enough for identifying the influential spreaders because the nearest neighbours of a well-connected node might have a low degree (Zhu, 2018). Therefore, it was reasonable to assume that the more connective nodes were the ones who had not only a high degree centrality but their neighbours were also well connected (Xu, Zhu, et al., 2020). Indeed, as the influential sub-groups, the major cities included in the maximum k-core region had well connected nearest neighbours to shape the transmission chains in the population networks.

3. Results

3.1. Analyzing the transmission patterns and outbreak sizes of COVID-19 in different transmission scenarios during phase one (1 January 2020–23 January 2020)

During phase one (1 January 2020–23 January 2020), before Wuhan’s quarantine, recent studies proposed that aggregate population outflow from Wuhan exhibited a highly and progressively correlation with infection prevalence across China (Guan et al., 2020). Therefore, the aggregate population outflow from the original epicentre can parsimoniously capture the early dynamics of COVID-19 transmission. In this paper, we set Wuhan as the baseline scenario and assumed other 42 major cities as the original epicentres, respectively. Then, we developed 42 transmission scenarios to show different distribution of infections with COVID-19 across China (Fig. 4). In each transmission scenario, the circles’ sizes represented the total aggregate population

outflows from the original epicentres, and the widths of the links represented the aggregate population outflows from the original epicentres to the destination cities. Comparing 42 transmission scenarios with the baseline scenario (Wuhan), we found that (1) the total aggregate population outflow from Wuhan (12.5) just ranked 20th among 43 major cities. The top five were Guangzhou (85.8-Level 9), Shanghai (73.5-Level 8), Shenzhen (72.9-Level 8), Dongguan (71.7-Level 8), and Beijing (69.9-Level 6), and four of these cities were first-tier cities. Most other 14 cities (i.e., Suzhou, Wenzhou, Nanjing, Wuxi, Ningbo, etc.), which also had higher total aggregate population outflows than Wuhan, were concentrated in the coastal areas of the Yangtze River Delta. By assuming a similar transmission mode as Wuhan, if COVID-19 emerged from one of these above cities, we inferred that more infections had been introduced into the destination cities over time. Substantial epidemic take-off in these destination cities would thus significantly contribute to COVID-19 transmission across China. Other 23 cities, which had lower total aggregate population outflows than Wuhan, were concentrated in the northeast (i.e., Harbin, Changchun), northwest (i.e., Urumqi, Xining, Yinchuan), or inland areas (i.e., Xi’An, Guiyang, Changsha) of China. (2) The number of destination cities in the baseline scenario (10 cities, seven provinces) just ranked 16th among 43 major cities. The top five were Nanchang (16 cities, 11 provinces), Jinan (13 cities, eight provinces), Qingdao (13 cities, ten provinces), Guiyang (13 cities, 11 provinces), and Beijing (12 cities, nine provinces). More destination cities implied that COVID-19 had transmitted more widely during the same observation period (1 January 2020–23 January 2020). (3) The standard deviation of the aggregate population outflow in the baseline scenario (0.48) just ranked 35th among 43 major cities, which implies that COVID-19 had spread uniformly across the destination cities in the baseline scenario (Fig. 5). Combined with the transmission routes of top 20 original epicentres (include Wuhan) (Table 1), some original epicentres had both the highest total aggregate population outflows and the highest standard deviation. For example, 79.34% of Foshan’s total aggregate population outflows were exported into Guangzhou. Therefore, Foshan-Guangzhou was the leading transmission chain in this transmission scenarios. By assuming the number of infections is uniformly distributed in the population outflows, the uneven geographical distribution of the aggregate population outflows caused COVID-19 to spread unevenly across the destination cities. If a public health crisis like COVID-19 emerges from one of these above original epicentres, we infers that more infections will be introduced into some specific destination cities with the higher proportion of the aggregate population outflows. These destination cities experience faster growth rates of COVID-19 during the same observation period (1 January 2020–23 January 2020). Therefore, when the second wave of COVID-19 comes, it could be valuable for public health experts and government officials to strictly impose quarantines in the major cities with the highest total aggregate population outflows and implement the targeted travel restrictions in the cities with the highest standard deviations. The combination of these interventions will be beneficial to interrupt COVID-19 transmission during the early phase of the epidemic.

3.2. Analyzing the importation risk of different major cities during phase two (24 January 2020–23 February 2020)

During phase two (24 January 2020–23 February 2020), after Wuhan’s quarantine, the confirmed COVID-19 cases had occurred outside Wuhan. There had been multiple epicentres in China, not just Wuhan. The newly imported cases in some major cities caused not only the local transmission but also the national COVID-19 outbreak due to the highly population flow among these cities. To capture the spreading process of COVID-19 across China accurately over time, we modeled the importation risk of 42 major cities during phase two (24 January 2020–23 February 2020). This model ranked 42 major cities according to their total importation risk values. Fig. 6 showed the top ten cities with the highest total importation risk values. A higher total importation

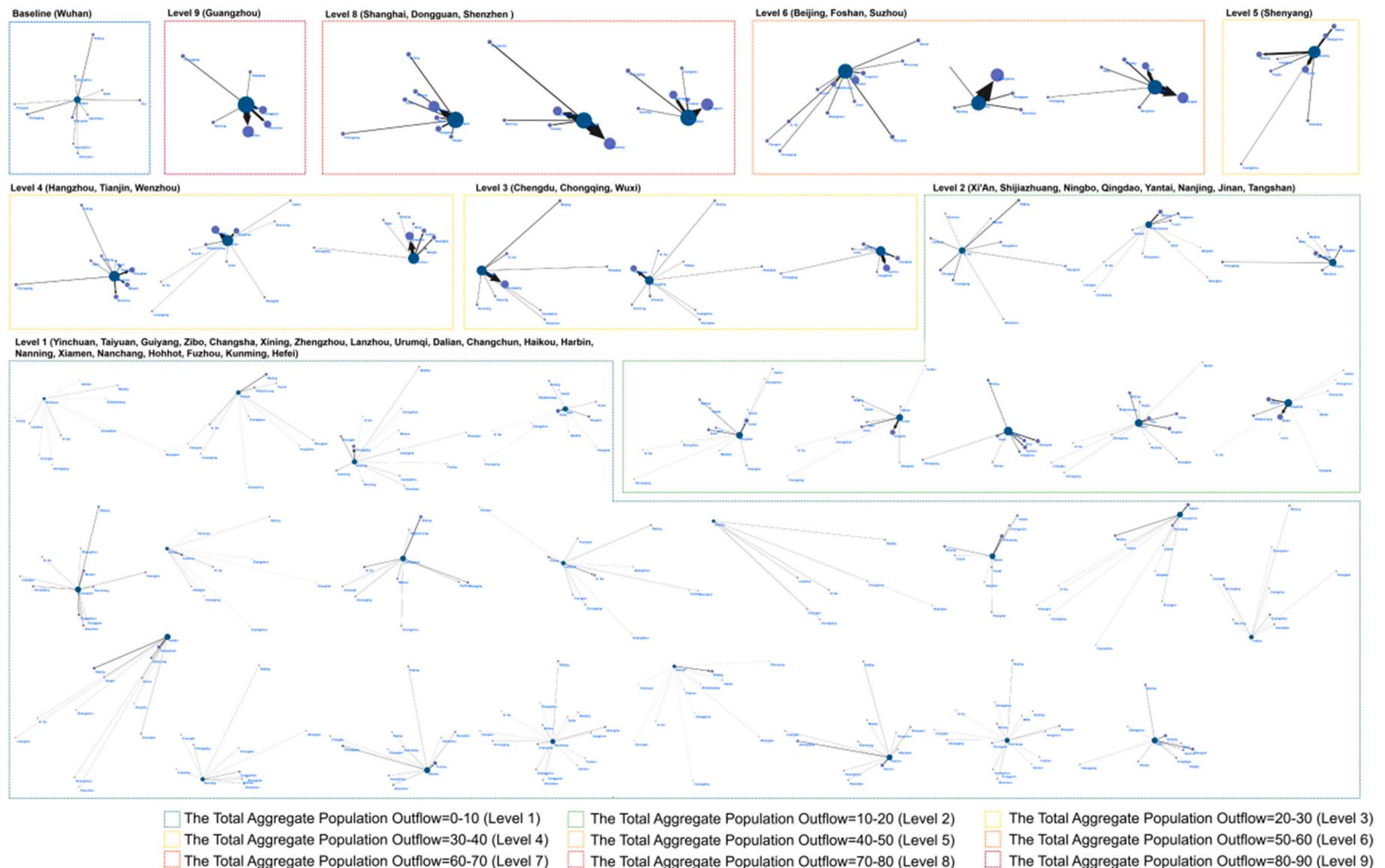


Fig. 4. Geographical distribution of the aggregate population outflows in different transmission scenarios during phase one.

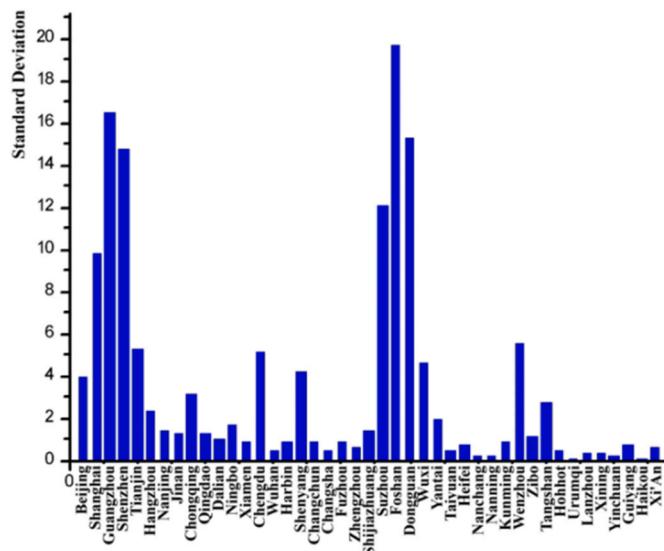


Fig. 5. The standard deviation of the aggregate population outflows of 43 major cities during phase one.

risk value predicted a higher level of COVID-19 transmission. Until 23 February 2020, the top three cities with the highest total importation risk values were Dongguan (522.04), Guangzhou (367.15), and Foshan (301.94). These three cities were located in the Pearl River Delta Economic Zone, which was one of the largest “mega-regions” across the world (Vidal, 2010). Other top ten cities included three first-tier cities (Beijing, Shanghai, Shenzhen), and two cities (Suzhou, Hangzhou) in the Yangtze River Delta Economic Zone. The standard deviation of the importation risk values was used to test the importation risks’ uneven distribution (Fig. 7). Combined with the importation risk distribution of the top ten cities (Table 2), some cities had both the highest total importation risk values and the highest standard deviations (such as Dongguan, Foshan, Guangzhou). For example, 71.49% of Dongguan’s total importation risk values were from Shenzhen. It represented Shenzhen was the major sources of Dongguan’s importation risk. Therefore, it might be possible to secure the containment of COVID-19 transmission by limiting the population flow from Shenzhen to Dongguan severely and immediately. Such successful, substantial, even draconian interventions can drastically reduce the importation risk of these major cities, which further controlled the early local transmission to not lead to a massive epidemic in these major cities. This model also yielded a dynamic performance metric to statistically derive the importation risk curves of 42 major cities, which enable to identify which cities perform better in controlling COVID-19 transmission. During phase two (24 January 2020–23 February 2020), among the top ten cities with the highest total importation risk values, four cities (Dongguan, Guangzhou, Foshan, Shenzhen) experienced the highest growth rates. They had maintained the highest growth rates until 23 February 2020, consistent with the highest total importation risk values. Then local epidemics probably grew exponentially in these cities. The growth rates of the other six cities had decreased gradually, especially some cities (Chengdu, Beijing, Tianjin) with the lowest growth rates have decreased to zero until 23 February 2020, which showed that more public health (non-pharmaceutical) interventions had been progressively and quickly implemented in these cities.

3.3. Analyzing the Spatio-temporal characteristics of the population flow and its influence on COVID-19 transmission during phase two (24 January 2020–23 February 2020)

To analyze the Spatio-temporal characteristics of the population flow and its influence on COVID-19 transmission during phase two (24

Table 1

Transmission routes of top 20 original epicentres during phase one.

| Top 20 original epicentres (total aggregate population outflow) | Destination cities (aggregate population outflow) |
|---|--|
| Baseline Wuhan (12.50) | Chongqing (1.97), Changsha (1.90), Beijing (1.84), Shanghai (1.36), Zhengzhou (1.07), Guangzhou (1.00), Shenzhen (0.96), Chengdu (0.84), Nanchang (0.82), Hefei (0.74) |
| Level 9 (total aggregate population outflow = 80–90) Guangzhou (85.83) | Foshan (45.42), Dongguan (16.74), Shenzhen (14.15), Chongqing (4.05), Nanning (2.87), Changsha (2.59) |
| Level 8 (total aggregate population outflow = 70–80) Shanghai (73.54) | Suzhou (33.09), Hangzhou (9.15), Wuxi (6.39), Nanjing (5.63), Beijing (5.50), Hefei (5.37), Ningbo (4.35), Chongqing (4.06) |
| Shenzhen (72.97) | Dongguan (39.90), Guangzhou (17.82), Chongqing (5.23), Foshan (4.50), Nanning (2.81), Changsha (2.70) |
| Dongguan (71.76) | Shenzhen (37.75), Guangzhou (21.94), Chongqing (4.56), Foshan (4.27), Nanning (3.24) |
| Level 6 (total aggregate population outflow = 50–60) Beijing (59.92) | Tianjin (16.87), Shijiazhuang (7.17), Tangshan (5.77), Shanghai (5.15), Harbin (3.84), Xi’An (3.53), Zhengzhou (3.44), Chengdu (3.03), Jinan (2.91), Taiyuan (2.84), Chongqing (2.68), Shenyang (2.66) |
| Foshan (59.14) | Guangzhou (46.92), Dongguan (3.57), Shenzhen (3.57), Nanning (2.64), Chongqing (2.44) |
| Suzhou (58.55) | Shanghai (32.50), Wuxi (14.05), Nanjing (4.69), Hangzhou (2.94), Hefei (2.70), Chongqing (1.67) |
| Level 5 (total aggregate population outflow = 40–50) Shenyang (41.64) | Dalian (13.05), Beijing (8.63), Changchun (6.68), Harbin (5.96), Tianjin (2.46), Shanghai (2.38), Guangzhou (1.25), Tangshan (1.23) |
| Level 4 (total aggregate population outflow = 30–40) Wenzhou (33.95) | Hangzhou (17.78), Suzhou (5.07), Shanghai (4.42), Wuxi (1.35), Ningbo (1.32), Hefei (1.10), Chongqing (1.06), Nanjing (0.93), Wenzhou (0.92) |
| Tianjin (33.68) | Beijing (16.67), Tangshan (8.52), Shijiazhuang (2.17), Harbin (1.28), Jinan (1.10), Shanghai (1.01), Xi’An (0.78), Shenyang (0.73), Taiyuan (0.71), Chongqing (0.69) |
| Hangzhou (30.80) | Shanghai (7.96), Ningbo (5.88), Wenzhou (5.47), Suzhou (2.46), Chongqing (2.45), Nanjing (2.00), Beijing (1.84), Hefei (1.66), Wuxi (1.08) |
| Level 3 (total aggregate population outflow = 20–30) Chengdu (26.13) | Chongqing (15.78), Beijing (2.37), Xi’An (2.15), Kunming (1.37), Shanghai (1.36), Guiyang (1.18), Guangzhou (0.99), Shenzhen (0.94) |
| Wuxi (25.01) | Suzhou (12.76), Shanghai (5.42), Nanjing (3.37), Hefei (1.25), Chongqing (1.12), Hangzhou (1.09) |
| Chongqing (21.10) | Chengdu (10.85), Guiyang (1.78), Beijing (1.47), Kunming (1.27), Shanghai (1.17), Xi’An (1.16), Guangzhou (0.91), Shenzhen (0.89), Wuhan (0.81) |
| Level 2 (total aggregate population outflow = 10–20) Nanjing (19.64) | Shanghai (4.46), Suzhou (3.99), Wuxi (3.30), Hefei (2.81), Beijing (1.92), Hangzhou (1.69), Chongqing (0.82), Wuhan (0.62) |
| Tangshan (16.10) | Tianjin (8.17), Beijing (5.08), Shijiazhuang (1.61), Shenyang (0.28), Harbin (0.26), Shanghai (0.16), Jinan (0.16), Changchun (0.15), Dalian (0.12), Xi’An (0.12) |
| Jinan (15.79) | (continued on next page) |

Table 1 (continued)

| Top 20 original epicentres (total aggregate population outflow) | Destination cities (aggregate population outflow) |
|---|---|
| Ningbo (14.25) | Zibo (3.99), Qingdao (3.27), Yantai (2.31), Beijing (2.28), Tianjin (0.74), Shanghai (0.69), Shijiazhuang (0.43), Nanjing (0.40), Zhengzhou (0.40), Chongqing (0.35), Xi'an (0.33), Harbin (0.32), Chengdu (0.30) Hangzhou (5.34), Shanghai (3.25), Wenzhou (1.61), Chongqing (1.59), Suzhou (1.08), Hefei (0.74), Nanjing (0.64) |
| Yantai (12.51) | Qingdao (6.72), Jinan (1.92), Beijing (0.93), Zibo (0.86), Dalian (0.66), Shanghai (0.41), Tianjin (0.30), Harbin (0.25), Xi'an (0.17), Zhengzhou (0.16), Chongqing (0.15) |

January 2020–23 February 2020), we constructed population flow networks (Fig. 8). 42 major cities were represented by nodes, whose sizes were proportional to their total importation risk values. Links among these major cities represented the population flow relationships. Noted that links were directional, and the widths of the links were proportional to their weights, which accounted for the importation risk values between these cities. Because the population flow evolved continuously, it increased the complexity of the network analysis. Here, according to the median incubation period of COVID-19 (four days) (Guan et al., 2020), we assumed that the cumulative number of the confirmed COVID-19 cases (per 1 million population) of 42 major cities had significantly changed every four days. Therefore, during phase two, we focused on the state of the population flow networks just at an interval of four days, which extracted the information for network analysis.

During phase two (24 January 2020–23 February 2020), network analysis of the population flow networks revealed their properties. Moreover, evolving in time, we tracked the evolution of these properties. At the system level, the density and the average clustering coefficient defined the population flow networks' topological characteristics in their unweighted versions. The densities (Fig. 9a) have increased from 0.30 to 0.44, which indicated that 14% of potential connections were actually realized until 23 February 2020. The average clustering coefficients (Fig. 9b) have increased from 0.54 to 3.34, which implied that the population flow networks had become significantly more clustered until 23 February 2020. Compared to the random network, these population flow networks exhibited a significantly higher average clustering coefficient, which were related to the more closely and tightly connected networks.

The k-core, combined with degree centrality, was used to recognize the influential sub-groups and the influential spreaders more accurately (Pei et al., 2014). As shown in Fig. 10, 42 major cities' connections in the population flow networks were visually displayed based on k-core regions. Eleven kinds of colours corresponded to 11 k-core partitions, and the red dotted circles corresponded to the top 5 cities with the highest degree centrality values. During phase two (24 January 2020–23 February 2020), in each population flow network, three or four connected sub-groups were identified with the k-core algorithm. And most of the major cities were included in the maximum k-core region, which had increased from 13-core to 19-core, and the number of these major cities had increased from 25 (59.5%) to 37 (88.1%). For example, Urumqi, Lanzhou, Xining, and Yinchuan had been included in the maximum k-core (16-core) from 28 January 2020. A larger k value implied a higher degree of connections in the sub-group. These growing and expanding population flow relationships among most major cities increased the transmission risk over time. Therefore, the sub-group with maximum k-core was defined as the influential sub-group in each population flow network. However, some major cities (such as Foshan included in 9-core region until 24 January 2020, Nanning and Dongguan

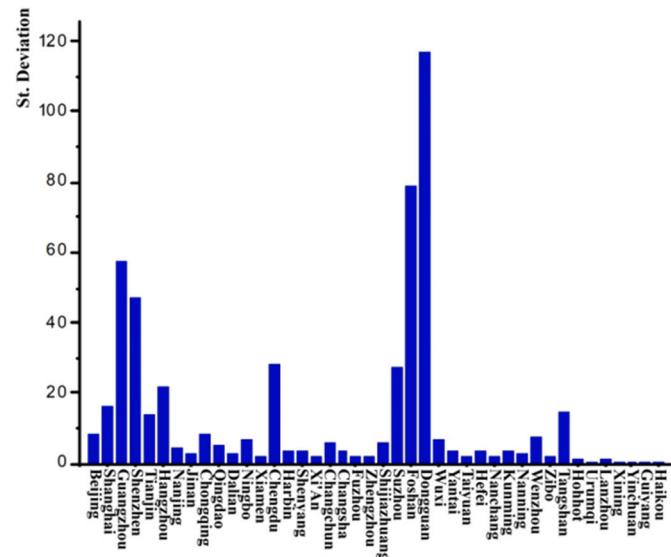


Fig. 7. The standard deviation of the importation risk values of 42 major cities during phase two.

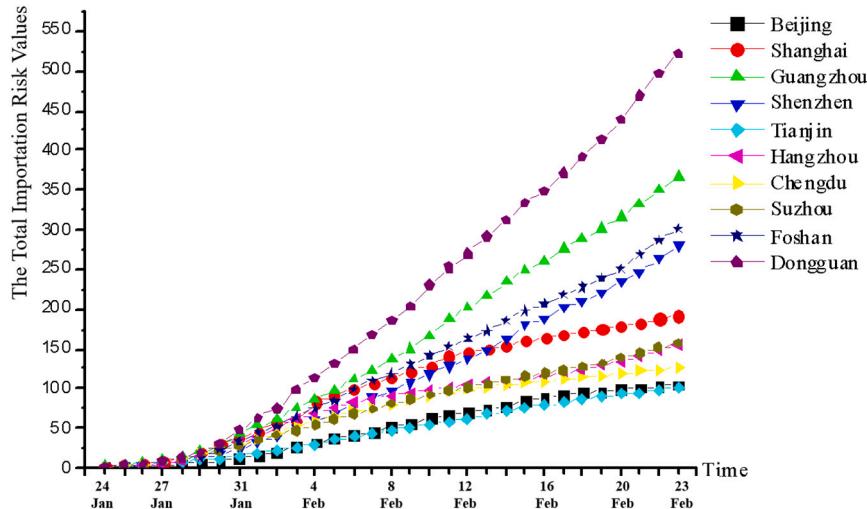


Fig. 6. The total importation risk values of top ten cities during phase two.

Table 2

The importation risk distribution of top ten cities during phase two.

| Top ten cities (total importation risk value) | Imported city (importation risk value) |
|---|--|
| Dongguan (522.04) | Shenzhen (373.21), Guangzhou (97.45), Chongqing (31.94), Foshan (12.54), Nanning (6.80), Beijing (0.04), Shanghai (0.03), Changsha (0.01), Chengdu (0.01) |
| Guangzhou (367.15) | Foshan (159.85), Shenzhen (105.70), Dongguan (80.50), Chongqing (18.16), Nanning (1.46), Changsha (0.87), Beijing (0.37), Shanghai (0.15), Chengdu (0.06), Hangzhou (0.02) |
| Foshan (301.94) | Guangzhou (242.04), Shenzhen (29.12), Dongguan (14.43), Chongqing (11.68), Nanning (4.50), Beijing (0.13), Shanghai (0.03), Changsha (0.01), Chengdu (0.01) |
| Shenzhen (280.34) | Dongguan (166.85), Guangzhou (62.44), Chongqing (30.60), Changsha (8.81), Foshan (6.83), Nanning (4.09), Beijing (0.32), Nanchang (0.20), Shanghai (0.11), Chengdu (0.05) |
| Shanghai (191.87) | Suzhou (57.47), Hefei (37.36), Chongqing (20.15), Wenzhou (22.29), Hangzhou (19.43), Ningbo (15.89), Wuxi (13.43), Nanjing (4.65), Beijing (0.62), Guangzhou (0.24) |
| Suzhou (157.07) | Shanghai (100.81), Wuxi (26.10), Hefei (13.88), Nanjing (10.72), Wenzhou (2.60), Chongqing (2.33), Hangzhou (0.39), Beijing (0.11), Ningbo (0.07), Guangzhou (0.02) |
| Hangzhou (155.97) | Wenzhou (91.44), Ningbo (26.13), Shanghai (19.09), Chongqing (8.71), Hefei (6.36), Suzhou (1.96), Beijing (1.18), Nanjing (0.68), Chengdu (0.13), Guangzhou (0.10) |
| Chengdu (126.42) | Chongqing (95.18), Shenzhen (7.75), Beijing (7.29), Xi'an (4.45), Shanghai (4.18), Guangzhou (2.94), Kunming (2.75), Guiyang (0.89), Wenzhou (0.79), Xining (0.16) |
| Beijing (105.08) | Tianjin (35.36), Harbin (20.48), Tangshan (15.65), Shijiazhuang (7.19), Chongqing (6.79), Xi'an (5.24), Changchun (2.65), Shenyang (2.63), Jinan (3.01), Taiyuan (1.57) |
| Tianjin (101.08) | Beijing (62.07), Tangshan (25.15), Harbin (6.88), Shijiazhuang (2.60), Jinan (1.36), Changchun (0.98), Shenyang (0.40), Chongqing (0.14), Yantai (0.10) |

included in 12-core region until 28 January 2020) included in the minimum k-core regions were periphery nodes, which had the lowest degree of connections with other major cities. These major cities may further decreased the transmission risk.

The cities with the highest degree centrality, marked in the red dotted circle, were defined as the influential spreaders in each population flow network. Due to the weighted population flow networks, the degree centrality did not only account for the number of links connected between the major cities, but also, how the importation risk values as the link weights were distributed. Therefore, two factors contributed to increasing the degree centrality: (i) more connections with other major cities and (ii) related higher importation risk values. During phase two (24 January 2020–23 February 2020), the degree centrality of these influential spreaders had increased significantly, such as the top three cities, Shenzhen (from 4.02 to 887.20), Guangzhou (from 3.56 to 809.34), and Dongguan (from 1.60 to 793.52).

Some cities (i.e., Guangzhou) had been the well-characterized influential spreaders during the whole phase two, which held a prominent structural advantage in facilitating continued COVID-19 transmission to other major cities. Some cities (i.e., Foshan) had become the potential influential spreaders during the late period of phase two, which played a pertinent role in generating new transmission chains to fuel and drive the COVID-19 transmission to more major cities. By contrast, the other cities (i.e., Chengdu) had not been the influential spreaders during the late period of phase two.

For analyzing the different impacts of these major cities on COVID-19 transmission, these three cities (Guangzhou, Foshan, Chengdu) were removed from the network, respectively. At the system level, the

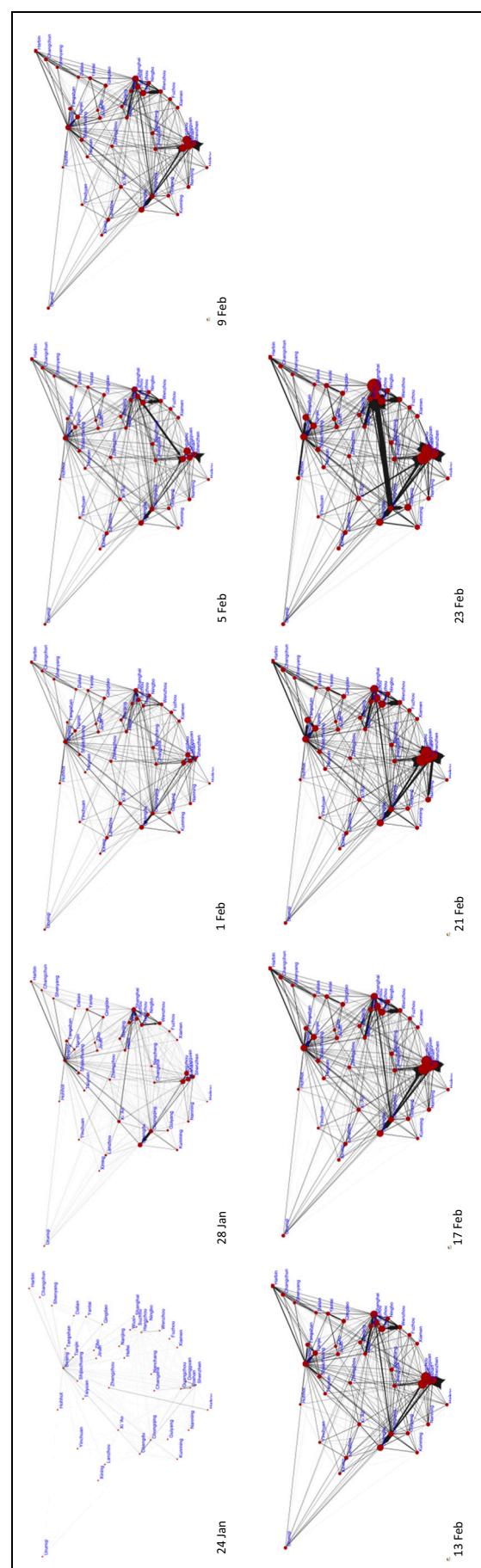


Fig. 8. The Spatio-temporal characteristics of population flow networks during phase two.

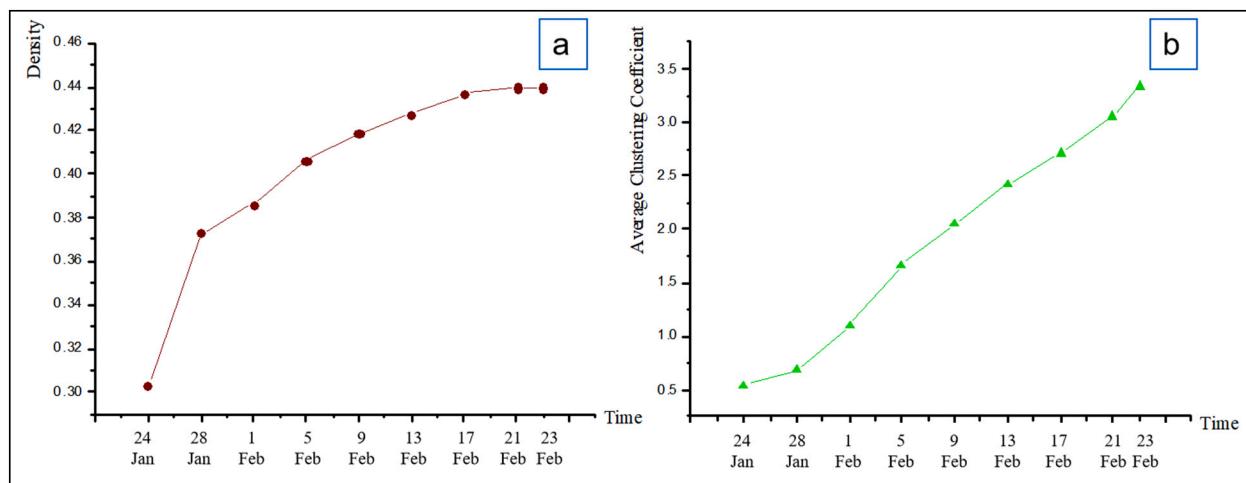


Fig. 9. The density and average clustering coefficient of the population flow networks during phase two.

population flow network's topological characteristics were most influenced by removing the Guangzhou. The density and the average clustering coefficient had averagely reduced by 2.55% and 8.98%, respectively. Therefore, targeting interventions or implementing a partial or total "travel restriction" in Guangzhou will significantly reduce the connections of the whole network during phase two, which can further strongly affect the dynamics of COVID-19 transmission. At the individual node level, by removing Guangzhou and Chengdu respectively, all connected sub-groups in the population flow networks were changed with the significantly decreased k values. The maximum k -core region included most major cities was decreased from 19-core to 17-core, which implied a highly decreased degree of connections in the influential sub-group. While removing Foshan, all connected sub-groups in this population flow network were not changed, which indicated the unchanged relations of COVID-19.

3.4. Analyzing the effectiveness of the travel restrictions among major cities

The transmission control (non-pharmaceutical) measures initiated during the spring migration, including the unprecedented Wuhan city travel ban and the level one national emergency response, had been proved to be strongly associated with, although not necessarily the cause of, a delay in epidemic growth and a reduction in confirmed case numbers (Tian et al., 2020). However, there has been no research to prove that the prohibiting travel among major cities (i.e., the suspended intercity public transport in 136 cities and the prohibited intercity travel in 219 cities) has reduced the numbers of confirmed COVID-19 cases. To test the effectiveness of travel restrictions among major cities in China, we estimated the total importation risk values of 42 major cities by using the population flow data during the same observation period (6 February 2019–8 March 2019) in 2019. By assuming the same importation ratio and the cumulative number of confirmed COVID-19 cases, our findings suggested that with the travel restrictions alone (without the Wuhan travel ban), the total importation risk values of 42 major cities had been decreased between 34.82% and 86.62%, with an average of 66.02% until the end of our observation period. Furthermore, most major cities' total importation risk values had been decreased up to 50% in the first ten days of the travel restrictions implemented, and the reduced rate and magnitude significantly had reached the peak in the first three days. Therefore, travel restrictions had sharply reduced the population flow among 42 major cities, which produced a much more significant effect on decreasing the total importation risk values of these major cities. Fig. 11 showed the percentage decrease of the total importation risk values (more than 70%) of the top ten cities with the

biggest percentage decrease during phase two (24 January 2020–23 February 2020). Due to the highly correlation between the total importation risk values and the cumulative number of confirmed COVID-19 cases, the sharp decrease of the total importation risk values represented the substantial reduction of the cumulative number of confirmed COVID-19 cases in the top ten cities, which explained the halted and reversed transmission of COVID-19.

4. Discussion and conclusion

In the early stage of COVID-19 outbreak, especially lacking the vaccine and the specific drug treatment for COVID-19, limiting the population flows among cities even countries by implementing the travel restrictions can drastically reduce initial imported seeding cases or even early local transmission (Bouchnita & Jebrane, 2020; Kraemer et al., 2020). However, precisely when and where the travel restrictions should be implemented is highly contextually specific. There is no one-size-fits-all set of prescriptive interventions being appropriate across all affected areas (Nishiura et al., 2020). Therefore, analyzing the dynamic of population flow and its influence on COVID-19 transmission could be crucial for public health response planning and control domestically and internationally. That is the critical contribution of our research. In summary, from 1 January 2020 to 23 January 2020 (phase one), for testing the contributions of different original epicentres to seeding epidemics elsewhere in China, we set Wuhan as the baseline scenario. We simulated 42 transmission scenarios by assuming different major cities as the original epicentres. From 24 January 2020 to 23 February 2020 (phase two), for testing the contribution of population flow among multiple epicentres to spreading COVID-19, we modeled the importation risk model to evaluate which cities had higher total importation risk. Then, the population flow networks weighted with the importation risk values were built to capture the distribution and evolution of COVID-19 transmission across China and identify influential sub-groups and influential spreaders. By removing different influential spreaders, we analyzed the different impacts of these influential spreaders on COVID-19 transmission. Targeting these vital sub-groups and major cities helps design strategies to either increase public health interventions' efficiency or hinder the diffusion of COVID-19.

The detailed findings of our research included as follows: first, during phase one, when exported infections from an original epicentre were dominant to determinate COVID-19 transmission pathways, we assumed 42 major cities as the original epicentres to simulate 42 transmission scenarios, separately. These transmission scenarios characterized the structure or relative distribution of infections across different geographical cities, driven by fundamentally total aggregate population

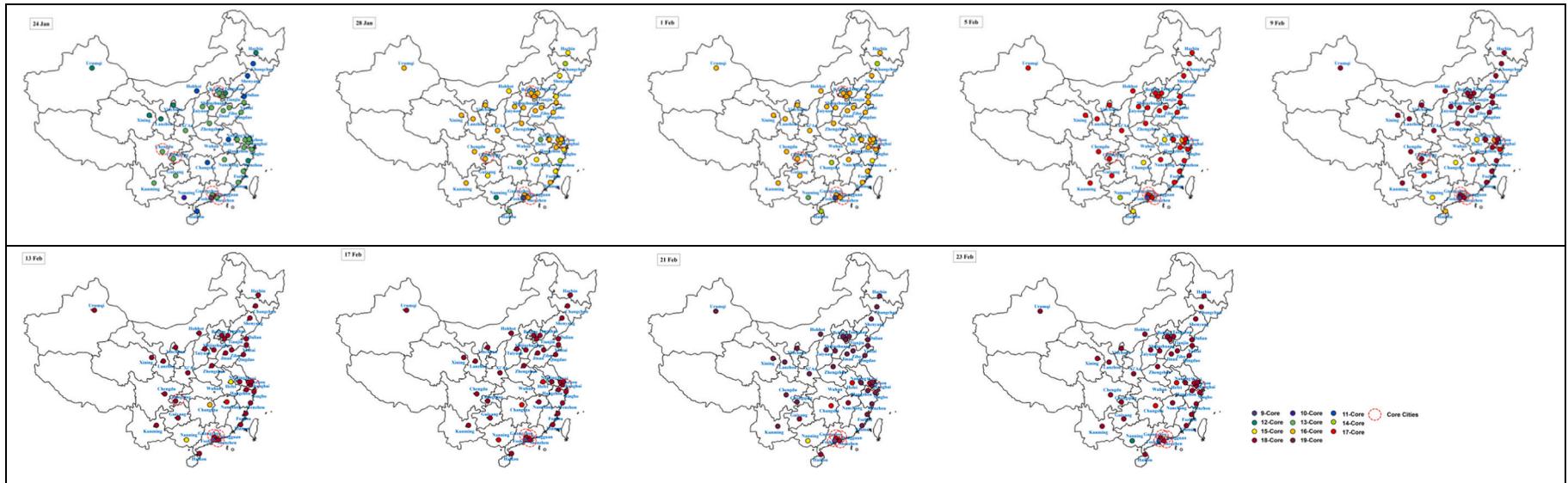


Fig. 10. Geographical distribution of the influential sub-groups and the influential spreaders in population flow networks during phase two transmission in the population flow network. The degree centrality of some major cities had also decreased significantly, and the newly influential spreaders had been identified by removing these three cities, respectively. For example, when removing Guangzhou, the degree centrality of the top 3 cities with the biggest percentage decreases were Foshan (80.85% decrease), Nanning (23.53% decrease), and Shenzhen (21.45% decrease). Hangzhou had become the newly influential spreader, Beijing and Chongqing had changed into the well-characterized influential spreaders, while Foshan had not been the potential influential spreaders. When removing Chengdu, the degree centrality of the top 3 cities were Chongqing (39.00% decrease), Kunming (20.67% decrease), and Xining (14.23% decrease). Hangzhou also have become the newly influential spreader, while Chongqing had not been the influential spreaders. When removing Foshan, the degree centrality of the top 3 cities were Guangzhou (38.73% decrease), Nanning (14.64% decrease), and Shenzhen (10.19% decrease). Replaced Guangzhou, Beijing had become the well-characterized influential spreaders.

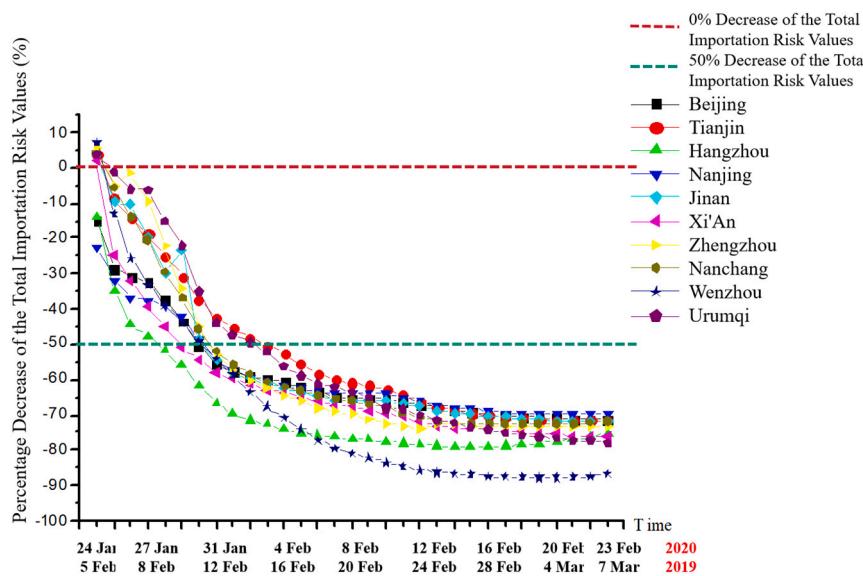


Fig. 11. The percentage decrease of the total importation risk values of the top ten cities during phase two (24 January 2020–23 February 2020).

outflows from these original epicentres. Compared with the baseline scenario (Wuhan, ranked 20th), the top five cities were global transport hubs with huge total aggregate population outflows (i.e., Guangzhou, Shanghai, Shenzhen, Dongguan, and Beijing). Four of these cities were first-tier cities. Other cities (i.e., Suzhou, Wenzhou, Nanjing, Wuxi, Ningbo, etc.), which also had higher total aggregate population outflows than Wuhan, were concentrated in the coastal areas of the Yangtze River Delta. If the COVID-19 outbreak occurs in one of these major cities, not Wuhan, there will be probably a substantially bigger scope and greater extent of transmission to other major cities. Furthermore, some original epicentres (i.e., Foshan) had both the highest total aggregate population outflows and the highest standard deviation, which made some transmission chains stronger (i.e., Foshan-Guangzhou). In the early stage of a new outbreak, this result highlights the strict control measures, including travel restrictions, that should be implemented in these certain transmission scenarios, which will provide the greatest benefit for mitigating the spread of COVID-19. Second, during phase two, when COVID-19 had already spread from Wuhan to other major cities within China, there were multiple epicentres forming multiple risk sources. We built the importation risk model to estimate the precise extent of the importation risk of 42 major cities. The total importation risk value was defined as the combination of the aggregate population inflow and the cumulative number of confirmed COVID-19 cases, explaining the spatial variation of the transmission rate. The correlation between the total importation risk and the cumulative number of confirmed COVID-19 cases provided support for our model's validity. Until 23 February 2020, three major cities (i.e., Dongguan, Guangzhou, Foshan) had the highest total importation risk values and growth rates. The highest number of infections imported into these major cities led to the exponential epidemic growth sharply. Furthermore, some cities' highest standard deviations indicated some specific core transmission chains (i.e., Dongguan-Shenzhen, Foshan-Guangzhou). Third, this research presented the first attempt to use network science to quantitatively characterize the structural evolution of the population flow networks, which gave a new perspective that analyzing and profiling the population flow network was fundamentally important for uncovering the dynamics of COVID-19 transmission. During phase two, we built the population flow networks as directed and weighted graphs. We investigated the characteristics of network properties both at the system level and the individual node level. At the system level, our results showed that the densities and the average clustering coefficients related to the population flow networks' topological characteristics had both increased

sharply. Therefore, until 23 February 2020, the population flow network's overall structure formed a more closely and tightly connected network. For the individual node level, our results showed that the major cities included in the maximum k-core region were highly mutually connected in each population flow network. These well connected major cities were defined as influential sub-groups. Until 23 February 2020, more major cities (from 25 to 37) had been included in the maximum k-core region, and the maximum k-core value had increased from 13-core to 19-core. The increasing size of the influential sub-group showed that more major cities were connected with stronger links, which increased their spreading capability of COVID-19. While other major cities not included in the influential sub-groups were defined as the peripheral nodes, which were less important for COVID-19 transmission. Then combined with degree centrality analysis, it was also worth recognizing the influential spreaders in these influential sub-groups. During the whole phase two, some cities (Shenzhen, Guangzhou) had always been the well-characterized influential spreaders, some major cities (Shanghai, Dongguan, Foshan) had become the influential potential spreaders later, and the other cities (Beijing, Chongqing, Chengdu) had not been the influential spreaders later. By removing these influential spreaders from the population flow network, respectively, we analyzed their different impacts on COVID-19 transmission. Removing Guangzhou can most influence the network's topological characteristics with the biggest reduced density and the average clustering coefficient. Moreover, the connections were uniformly distributed in the population flow network. Removing Guangzhou and Chengdu can significantly change the structure of all sub-groups with the decreased k values. The dominant roles of other influential spreaders had been changed by removing these influential spreaders. When Guangzhou and Chengdu were removed, respectively, some major cities (i.e., Foshan, Chongqing) had not been the potential influential spreaders due to the biggest percentage decrease of degree centrality, some major cities (i.e., Hangzhou) had become the newly influential spreader. Therefore, infection prevention and control policies aimed at these targeted influential sub-groups and influential spreaders can effectively limit the spread of COVID-19.

Since 23 January 2020, strict travel restrictions were initiated between major cities or provinces across China. Combined with other public health interventions, including suspending intracity public transport, closing entertainment venues, banning public gatherings, and isolating suspected and confirmed patients, there was no evidence to prove the travel restrictions' effectiveness on reducing the numbers of

confirmed COVID-19 cases. Using the population flow data during the same observation period (6 February 2019–8 March 2019) in 2019, our results showed that the total importation risk values of 42 major cities had been averagely reduced 66.02%, just due to the travel restrictions alone. Therefore, when there are multiple epicentres across a country, prohibiting travel in targeted influential sub-groups and influential spreaders will be the most effective control measure to contain COVID-19 transmission strongly.

Despite the comprehensive findings, this research has several limitations. First, due to the current limited open access availability of data, our observation period was just limited to the first 54 days of COVID-19 in China. The observation period's length could be adapted to capture the “epidemic peak” of COVID-19, which may change from city to city or even country to country. Second, the importation risk model was built based on the population flow networks. For simplicity, we assumed that the importation risk value was only defined as the combination of the aggregate population inflow and the cumulative number of confirmed COVID-19 cases. It would be beneficial to identify other possible parameters, such as travel passengers' different demographic characteristics and travel behaviour's different patterns associated with different travel restrictions, which are crucial to affect the importation risk value. A further limitation is that, a wide range of other public health interventions had been implemented together in addition to travel restrictions. Owing to lack of valid and reliable data, we could not separate their individual effects, or conduct a comparative effectiveness analysis of specific interventions in this research. Therefore, in this research, we only provide a rapid and relatively crude assessment of travel restrictions at a relatively early stage of the COVID-19 outbreak. During the second wave of COVID-19, the comprehensive prevention interventions (i.e., full-covered COVID-19 testing, tracing and isolating all close contacts) were implemented in most major cities in China. Therefore, as the COVID-19 pandemic continues to evolve, this epidemic was effectively contained in some cities which were located in the coastal areas of the Yangtze River Delta, though they had higher total aggregate population outflows or higher importation risk values. Future research will be needed to examine which interventions should be implemented first as the epidemic curve grew or which interventions should be lifted first as the epidemic curve flattened.

CRediT authorship contribution statement

Jie Liu: Methodology, Software, Visualization, Writing-Original Draft Preparation.

Jingyu Hao: Investigation, Validation.

Yuyu Sun: Data Curation.

Zhenwu Shi: Conceptualization, Supervision, Writing-Reviewing and Editing.

Declaration of competing interest

None.

Acknowledgments

Funding by the National Social Science Fund of China (No. 20CGL051) were gratefully acknowledged.

References

- Baker, T. L., Greiner, J. V., Maxwell-Schmidt, E., Lamothe, P. H., & Vesonder, M. (2020). Guidelines for frontline health care staff safety for COVID-19. *Journal of Primary Care & Community Health*. <https://doi.org/10.1177/2150132720938046>.
- Bherwani, H., Anjum, S., Kumar, S., Gautam, S., Gupta, A., Kumbhare, H., Anshul, A., & Kumar, R. (2020). Understanding COVID-19 transmission through Bayesian probabilistic modeling and GIS-based Voronoi approach: a policy perspective. *Environment, Development and Sustainability*. <https://doi.org/10.1007/s10668-020-00849-0>.
- Bouchnita, A., & Jebrane, A. (2020). A multi-scale model quantifies the impact of limited movement of the population and mandatory wearing of face masks in containing the COVID-19 epidemic in Morocco. *Mathematical Modelling of Natural Phenomena*, 15, 31.
- Burleson-Lesser, K., Morone, F., Tomassone, M. S., & Makse, H. A. (2020). K-core robustness in ecological and financial networks. *Scientific Reports*, 10, 3357–3370.
- Chen, S., Yang, J., & Yang, W. (2020). COVID-19 control in China during mass population movements at New Year. *Lancet*, 395(10226), 764–766.
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., ... Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368, 395–400.
- Ghalmane, Z., Cherif, C., Cherif, H., & Hassouni, M. E. (2019). Centrality in complex networks with overlapping community structure. *Scientific Reports*, 9, 10133–10161.
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 26, 855–860.
- Giustolisi, O., Ridolfi, L., & Simone, A. (2020). Embedding the intrinsic relevance of vertices in network analysis: The case of centrality metrics. *Scientific Reports*, 10, 3297–3307.
- González, M., Hidalgo, C., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, 453, 779–782.
- Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., ... Zhong, N. (2020). Clinical characteristics of coronavirus disease 2019 in China. *The New England Journal of Medicine*, 382, 1708–1720.
- He, S., Peng, Y., & Sun, K. (2020). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics*, 101, 1667–1680.
- Hou, C., Chen, J., Zhou, Y., Hua, L., Yuan, J., He, S., ... Jia, E. (2020). The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis. *Journal of Medical Virology*, 92(7), 841–848.
- Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 582, 389–394.
- Kraemer, M. U. G., Yang, C. H., Gutierrez, B., Wu, C., Klein, B., Pigott, D. M., ... Scarpino, S. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368, 493–497.
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., & Eggo, R. M. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases*, 20, 553–558.
- Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, 55(3), 105924.
- Liu, Y., Tang, M., Zhou, T., & Do, Y. (2015). Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. *Scientific Reports*, 5, 9602–9609.
- National Bureau of Statistics of China. (2019). China Statistical Yearbook. <http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm>.
- National Health Commission of the People's Republic of China, & Traditional Chinese Medicine of the People's Republic of China. (2020). Guidelines for the Diagnosis and Treatment of Coronavirus Disease 2019 (Trial Version 7). <http://www.nhc.gov.cn/yzyg/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>.
- Nishiura, H., Kobayashi, T., Yang, Y., Hayashi, K., Miyama, T., Kinoshita, R., ... Akhmetzhanov, A. R. (2020). The rate of underascertainment of novel coronavirus (2019-nCoV) infection: Estimation using Japanese passengers data on evacuation flights. *Journal of Clinical Medicine*, 9, 419–421.
- Pei, S., Muchnik, L., Andrade, J. S., Jr., Zheng, Z., & Makse, H. A. (2014). Searching for superspreaders of information in real-world social media. *Scientific Reports*, 4, 5547–5558.
- Perkins, T. A., Scott, T. W., Le Menach, A., & Smith, D. L. (2013). Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLOS Computational Biology*, 9(12), Article e1003327.
- Shi, Q., Dorling, D., & Cao, G. (2020). Changes in population movement make COVID-19 spread differently from SARS. *Social Science & Medicine*, 255, 113036.
- Song, C., Koren, T., Wang, P., & Barabási, A. L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, 6, 818–823.
- Tian, H., Liu, Y., Li, Y., Wu, C. H., Chen, B., Kraemer, M. U. G., ... Dye, C. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, 368, 638–642.
- Tiwari, S., Kumar Jha, S., & Singh, A. (2020). Quantification of node importance in rain gauge network: Influence of temporal resolution and rain gauge density. *Scientific Reports*, 10, 9761–9777.
- Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Ramirez-Paredes, J., Halsey, E. S., Kochel, T. J., Scott, T. W., & T.W., & Kitron, U. (2013). Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *Plos One*, 88(4), Article e58802.
- Vidal, J. (2010). UN Report: World's Biggest Cities Merging into 'Mega-Regions'. Available online guardian.co.uk. (Accessed 13 March 2010).
- Wiwanikit, V., & Joob, B. (2020). Density of COVID-19 and mass population movement during long holiday: Simulation comparing between using holiday postponement and no holiday postponement. *Journal of Research in Medical Sciences*, 25, 55.
- World Health Organization. (2020). WHO Coronavirus Disease (COVID-19) Dashboard (Data last updated: 2020/7/29, 8:00am) <https://covid19.who.int/>.
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet*, 395(10225), 689–697.

- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., & Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China Life Sciences*, 63(3), 457–460.
- Xu, X., Zhu, C., Wang, Q., Zhu, X., & Zhou, Y. (2020). Identifying vital nodes in complex networks by adjacency information entropy. *Scientific Reports*, 10, 2691–2701.
- Yang, C. H., & Jung, H. (2020). Topological dynamics of the 2015 South Korea MERS-CoV spread-on-contact networks. *Scientific Reports*, 10, 4327–4337.
- Yang, Z., Zeng, Z., Wang, K., Wong, S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., ... He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 12(3), 165–174.
- Yue, X. G., Shao, X. F., Li, R. Y. M., Crabbe, M. J. C., Mi, L., Hu, S., ... Liang, G. (2020). Risk management analysis for novel coronavirus in Wuhan, China. *Journal of Risk and Financial Management*, 13, 22–27.
- Zhu, F. (2018). Improved collective influence of finding most influential nodes based on disjoint-set reinsertion. *Scientific Reports*, 8, 14503–14514.

GitHub

GIT CHEAT SHEET

Git is the free and open source distributed version control system that's responsible for everything GitHub related that happens locally on your computer. This cheat sheet features the most important and commonly used Git commands for easy reference.

INSTALLATION & GUIs

With platform specific installers for Git, GitHub also provides the ease of staying up-to-date with the latest releases of the command line tool while providing a graphical user interface for day-to-day interaction, review, and repository synchronization.

GitHub for Windows

<https://windows.github.com>

GitHub for Mac

<https://mac.github.com>

For Linux and Solaris platforms, the latest release is available on the official Git web site.

Git for All Platforms

<http://git-scm.com>

SETUP

Configuring user information used across all local repositories

git config --global user.name "[firstname lastname]"

set a name that is identifiable for credit when reviewing version history

git config --global user.email "[valid-email]"

set an email address that will be associated with each history marker

git config --global color.ui auto

set automatic command line coloring for Git for easy reviewing

SETUP & INIT

Configuring user information, initializing and cloning repositories

git init

initialize an existing directory as a Git repository

git clone [url]

retrieve an entire repository from a hosted location via URL

STAGE & SNAPSHOT

Working with snapshots and the Git staging area

git status

show modified files in working directory, staged for your next commit

git add [file]

add a file as it looks now to your next commit (stage)

git reset [file]

unstage a file while retaining the changes in working directory

git diff

diff of what is changed but not staged

git diff --staged

diff of what is staged but not yet committed

git commit -m "[descriptive message]"

commit your staged content as a new commit snapshot

BRANCH & MERGE

Isolating work in branches, changing context, and integrating changes

git branch

list your branches. a * will appear next to the currently active branch

git branch [branch-name]

create a new branch at the current commit

git checkout

switch to another branch and check it out into your working directory

git merge [branch]

merge the specified branch's history into the current one

git log

show all commits in the current branch's history



INSPECT & COMPARE

Examining logs, diffs and object information

`git log`

show the commit history for the currently active branch

`git log branchB..branchA`

show the commits on branchA that are not on branchB

`git log --follow [file]`

show the commits that changed file, even across renames

`git diff branchB...branchA`

show the diff of what is in branchA that is not in branchB

`git show [SHA]`

show any object in Git in human-readable format

SHARE & UPDATE

Retrieving updates from another repository and updating local repos

`git remote add [alias] [url]`

add a git URL as an alias

`git fetch [alias]`

fetch down all the branches from that Git remote

`git merge [alias]/[branch]`

merge a remote branch into your current branch to bring it up to date

`git push [alias] [branch]`

Transmit local branch commits to the remote repository branch

`git pull`

fetch and merge any commits from the tracking remote branch

TRACKING PATH CHANGES

Versioning file removes and path changes

`git rm [file]`

delete the file from project and stage the removal for commit

`git mv [existing-path] [new-path]`

change an existing file path and stage the move

`git log --stat -M`

show all commit logs with indication of any paths that moved

REWRITE HISTORY

Rewriting branches, updating commits and clearing history

`git rebase [branch]`

apply any commits of current branch ahead of specified one

`git reset --hard [commit]`

clear staging area, rewrite working tree from specified commit

IGNORING PATTERNS

Preventing unintentional staging or committing of files

`logs/ *.notes pattern*/`

Save a file with desired patterns as .gitignore with either direct string matches or wildcard globs.

`git config --global core.excludesfile [file]`

system wide ignore pattern for all local repositories

TEMPORARY COMMITS

Temporarily store modified, tracked files in order to change branches

`git stash`

Save modified and staged changes

`git stash list`

list stack-order of stashed file changes

`git stash pop`

write working from top of stash stack

`git stash drop`

discard the changes from top of stash stack

GitHub Education

Teach and learn better, together. GitHub is free for students and teachers. Discounts available for other educational uses.

✉ education@github.com

☞ education.github.com

The Employees' Pension Scheme,



FORM 2

EMPLOYEES' PROVIDENT FUNDS SCHEME, 1952

[Paragraphs 33 AND 61(1)]

EMPLOYEES' PENSION SCHEME, 1995

[Paragraphs 18]

DECLARATION AND NOMINATION FORM UNDER THE EMPLOYEES' PROVIDENT FUNDS SCHEME AND EMPLOYEES' PENSION SCHEME

1. UAN : 101703664453

2. Name (in block letters) : ANUJ VERMA

3. Father's / Husband's Name : LATE ASHOK VERMA

4. Date of Birth : 01/01/1998

5. Gender : MALE

6. Marital Status : UN-MARRIED



7. Address (Permanent) : House no 4 Block no 31 , Bhrigu Ashram, Ballia, BALLIA, UTTAR PRADESH, 277001

8. Address (Temporary) : House no 4 Block no 31 , Bhrigu Ashram, Ballia, BALLIA, UTTAR PRADESH, 277001

9. (A) Date of Joining of EPF Scheme, : 14/06/2021

(B) Date of Joining of FPS Scheme, : --

(C) Date of Joining of EPS Scheme, : --

PART A (EPF)

I hereby nominate person(s) / cancel the nomination made by me previously and nominate the person (s) mentioned below to receive the amount standing to my credit in the Employees' Provident Fund, in the

| Name of the Nominee / Nominees (1) | Address (2) | Nominees's relationship with the Member (3) | Date of Birth (4) | Total amount or share of accumulations in Provident Fund to be paid to each nominee (5) | If the Nominee is minor, name and relationship and address of the guardian who may receive the amount during the minority of nominee (6) |
|---|---|--|----------------------|--|---|
|  GEETA VERMA AADHAAR: XXXX XXXX 4356 Bank A/c: NOT AVAILABLE | House no 4 Block no 31 , Bhrigu Ashram, Ballia, BALLIA, UTTAR PRADESH, 277001 | Dependent Mother | 01/01/1970 | 100% | -- |

1. *Certified that I have no family as defined in Para 2(g) of the Employees' Provident Funds Scheme, 1952 and should I acquire a family hereafter the above nomination should be deemed as cancelled.

2. *Certified that my father/mother is/are dependant upon me.

(Signature is not required as the document is to be digitally signed)

The Employees' Pension Scheme,

PART B (EPS)
(Para 18)

I hereby furnish below particulars of the members of my family who would be eligible to receive widow / children pension in the event of my death:

| Sl. No. (1) | Name of the Family members (2) | Address (3) | Date of Birth (4) | Relationship with the member (5) |
|----------------|-----------------------------------|----------------|----------------------|-------------------------------------|
| 1 -- | <input type="text"/> | -- | -- | -- |

*Certified that I have no family as defined in Para 2(vii) of the Employees' Pension Scheme, 1995 and should I acquire a family hereafter I shall furnish particulars thereon in the above form.

I hereby nominate the following person for receiving the monthly widow pension admissible under para 16(2)(a)(i) and (ii) in the event of my death without leaving any eligible family member for receiving

| Name and Address of the Nominee | Date of Birth | Relationship with the member |
|--|---------------|------------------------------|
|  GEETA VERMA AADHAAR: XXXX XXXX 4356 Bank A/c: NOT AVAILABLE | 01/01/1970 | DEPENDENT MOTHE |

Date : 14-Feb-2022

(Signature is not required as the document is to be digitally signed)

Network Biology

Coding Assignment 1

Note: You are responsible for the backup of data as well as results, which will be used for evaluation.

Code for the following tasks. Do not use existing libraries (such as NetworkX/igraph) for network analysis. You need to submit the Jupyter Notebook, its PDF as well we the data files for evaluation.

1. Draw a simple network (of ~ 10 nodes and ~ 20 edges) of each of the following classes: undirected & unweighted (UD-UW); undirected and weighted (UD-W); directed and unweighted (D-UW); directed and weighted (D-W). Store these networks using (i) adjacency matrix and (ii) edge list representation. [2]
2. Write a program to compute the degree (for undirected graphs) and in/out-degree (for directed graphs) of each node. [3]
3. Starting from the adjacency matrix (A) of the UD-UW graph, compute A^2 . What information, if any, does it hold in the context of the graph you have drawn in response to exercise 1? What about higher-order matrix multiplications of A ? [3]
4. Download data of networks available from Mark Newman's page (<http://www-personal.umich.edu/~mejn/netdata/>). Read the data, and measure the degree of nodes (using the program written in exercise 2). Compute and plot the 'degree distribution'. [2]
5. Write a program to create a random graph/network (UD-UW) of n nodes and e edges, and to compute its 'degree distribution'. [5]
6. For real-world networks used in exercise 4, create 100 instances of random controls (by preserving the number of nodes and edges). Compute and plot its average degree distribution. Comment on differences in degree distributions observed in real-world networks and their random counterparts. [5]
7. Read the data of real-world networks into Cytoscape. Use Cytoscape to compute their degree distributions. Visualize the networks and export their images in JPG. [2]
8. Write a program to compute the **clustering coefficient** of node (C_i) and average clustering coefficient of the network (C) starting from the adjacency matrix of an undirected graph. [5]
9. Write a program to compute the **shortest path length** between a pair of nodes (L_{ij}) and the characteristic path length of the network (L) starting from the adjacency matrix. (Hint: Look at well-known algorithms such as that by Dijkstra.) [5]
10. Download the HPRD data. Compute the number of nodes and edges. Plot the degree distribution of the protein-protein interaction network on a log-log scale. [3]