Infectious Disease Modelling 8 (2023) 1117–1126

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

The spatiotemporal analysis of the population migration network in China, 2021

Wenjie Li [a], Ye Yao [a, b, *]

[a] Department of Biostatistics, School of Public Health, Fudan University, Shanghai, China
[b] Key Laboratory of Public Health Safety of Ministry of Education, Fudan University, Shanghai, China

ARTICLE INFO

Article history:
Received 14 November 2022
Received in revised form 6 October 2023
Accepted 10 October 2023
Available online 11 October 2023
Handling Editor: Dr Yiming Shao

Keywords:
K-shell decomposition
Louvain algorithm
Population mobility
Infectious disease
Network analysis

ABSTRACT

Population migration is a critical component of large-scale spatiotemporal models of infectious disease transmission. Identifying the most influential spreaders in networks is vital to controlling and understanding the spreading process of infectious diseases. We used Baidu Migration data for the whole year of 2021 to build mobility networks. The nodes of the network represent cities, and the edges represent the population flow between cities. By applying the k-shell decomposition and the Louvain algorithm, we could get the k-shell values for each city and community partition. Then, we identified the most efficient nodes or pathways in a complex network by generating random networks. Furthermore, we analyzed the eigenvalue of the migration matrix to find the nodes that have the most impact on the network. We also found the consistency between k-shell value and eigenvalue through Kendall's τ test. The main result is that in Spring Festival and National Day, the network is at higher risk of an infectious disease outbreak and the Yangtze River Delta is at the highest risk of an epidemic all year around. Shanghai is the most significant node in both k-shell value and eigenvalue analysis. The spatiotemporal property of the network should be taken into account to model the transmission of infectious diseases more accurately.

1. Introduction

Since the beginning of the 21st century, humans have experienced the Severe Acute Respiratory Syndrome (SARS) outbreak in 2003(Shaw, 2006), the H1N1 Influenza outbreak in 2009(Louie et al., 2010), the outbreak of the Ebola virus in western Africa from 2014 to 2016(Henao-Restrepo et al., 2016), the explosion of Middle East Respiratory Syndrome (MERS) in 2015(Müller et al., 2015), and the pandemic of Corona Virus Disease 2019 (COVID-19) from the end of 2019 to the present(Velavan & Meyer, 2020; Wu et al., 2020) and so on. As of November 4, 2022, there have been 628,694,934 confirmed cases of COVID-19, including 6,576,088 deaths, reported to WHO. The outbreak of these infectious diseases has seriously disrupted people's regular life, endangered people's lives and property, and at the same time has had a serious impact on the country's economic development(Zhang et al., 2022).

* Corresponding author. Department of Biostatistics, School of Public Health, Fudan University, No. 130 Dongan Road, Shanghai, 200032, China.
E-mail address: yyao@fudan.edu.cn (Y. Yao).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2023.10.003

The main transmission routes of infectious diseases such as COVID-19 are respiratory droplets and contact transmission. The spread of the virus across the country and even the world is mainly due to the long-distance migration of its hosts. For example, the global spread of SARS in 2003(Colizza et al., 2006). When the virus follows the host from one place to another, the newly infected area may become another source and contaminate other areas. Research pointed out that the epidemic of SARS in Hong Kong, China began to be controlled in early April 2003, mainly because people took various measures to effectively reduce the chance of contact between infected and healthy people(Dye & Gay, 2003). It can also be found that similar to the global spread of SARS, the migration behavior of individuals also plays a crucial role in the influenza A (H1N1) epidemic(Khan et al., 2009; Mukherjee et al., 2010). These facts inspire us to consider the mobility of individuals to reflect the real world more accurately when we simulate the spread of infectious diseases.

In the study of the spatiotemporal spread of infectious diseases, a city can be regarded as a node, and population flow can be viewed as a connection. This way, we can construct the complex network of the spatiotemporal spread of infectious diseases based on real-world population flow data. As a mature scientific research method, complex network analysis can be used to explore the inherent characteristics and propagation mechanism of spatiotemporal diffusion. Vespignani's group used global airline data to build an airline transportation network. They tried to predict pandemics and simulate the risk of spreading infectious diseases from a single city to other cities worldwide (Colizza et al., 2006). Xin Lu's group used mobile phone signalling data to simulate the migration patterns of people in different periods in China. Using the Louvain algorithm, cities were divided into ten communities, and the relevant results could be applied to the prevention and control of the diffusion of infectious diseases(Tan et al., 2021). Olha Buchel's group applied the Louvain method to the networks built based on individuals' movements and found that restricting commutes between low and high-risk patches could control the spread of coronavirus across various areas(Buchel et al., 2021).

The centrality of nodes is of great significance in the analysis of disease spread. Many studies adopted nodal centrality analysis to predict(Bucur & Holme, 2020) and control the spread of disease(Chaharborj et al., 2022; Christley et al., 2005) as well as target immunization strategies(Clusella et al., 2016; Wei et al., 2022). The most common metrics used to describe the importance of nodes in complex network analysis are degree centrality, betweenness centrality(Freeman, 1977), and closeness centrality(Krackhardt, 1990). Besides, several algorithms have been introduced, including the collective influence (CI), the algorithms based on random walk, belief propagation-guided decimation (BPD) algorithm(Zhao et al., 2020). Many studies have shown that k-shell value is more suitable for portraying node importance(Kitsak et al., 2010). The k-shell value has the greatest impact on information spreading(Zhu & Zhang, 2019) and is the most correlated to epidemic spreading(de Arruda et al., 2014). In this work, we analyzed the spatiotemporal features of the population mobility network to provide parameters for simulating the spread of infectious diseases. We used Baidu Migration data for the whole year of 2021 to build mobility networks. By applying the k-shell decomposition and Louvain algorithm, we could get k-shell values for every city and community partition. Then, we identified the most efficient nodes or pathways in a complex network by generating random networks. Furthermore, we analyzed the eigenvalue of the migration matrix to find significant nodes that have the greatest nationwide impact on the propagation of infectious disease. It's shown that k-shell value is consistent with the eigenvalue through Kendall test. The more central the node's position is, the greater effect it will have on the network.

2. Material and methods

2.1. Data

2.1.1. Baidu Migration data

The data was obtained from the Baidu Migration website (http://qianxi.baidu.com/). Baidu Migration analyzes the location information collected by the Baidu Map LBS open platform to obtain population flow trajectories, updates the population flow between 366 cities, and displays the percentage of the population moving from one city to another. Percentage x_{ij} represents the proportion of the outflow population from city i to city j to the total outflow population of city i.

We collected population flow between all 366 cities for the whole year of 2021. Compared with 2020 and 2022, the data of 2021 could generally represent population mobility because there was no influential lockdown in China this year. Findings have shown that there was a 30.1% mobility reduction due to lockdowns(Joshi & Musalem, 2021). The nationwide lockdown in 2020 and the lockdown in Shanghai in 2022 might significantly impact human mobility.

The data was divided into four periods according to the official: Spring Festival (January 28-March 8), National Day (October 1-7), weekends, and workdays. With the percentage of the population moving from one city to another, we could get a migration matrix for each day. Then, get the mean migration matrix averaged from the everyday migration matrix for each period. In order to avoid the interference of extremely low edges, we eliminated edges with percentages lower than 1 percent. Finally, we have got more than 4,000 edges in each period.

2.1.2. Real epidemic data

We collected origin cities that had reported new COVID-19 cases from March 1, 2022, to March

31, 2022. We have found out the number of cities that were affected (news or reports saying that the confirmed case in the affected city was from the origin city) within 14 days of the first confirmed case in the origin city. Jilin (first case confirmed on

March 2, 2022) could spread to 7 cities within 14 days. Lanzhou (first case confirmed on March 6, 2022) could spread to 4 cities, while Shanghai (first case confirmed on March 1, 2022) could spread to 11 cities, and Quanzhou (first case confirmed on March 14, 2022) could spread to 8 cities.

2.2. Methods

With a migration matrix, we could construct a mobility network, 366 nodes representing cities and weight of edge representing the percentage of the population moving from one city to another. As for the parts of k-shell decomposition and generating random networks, we constructed an unweighted, undirected network for each period based on the mean migration matrix. While in the Louvain algorithm and eigenvalue analysis, we constructed a weighted directed network.

2.2.1. K-shell decomposition

K-shell decomposition(Kitsak et al., 2010) is based on degree centrality and can divide nodes in the network into different layers according to their importance. First, find all nodes with degree $k = 1$ in the network. From the point of degree centrality, these nodes are the least important in the network. Remove all these nodes and their connected edges. At this point, some new nodes with a degree $k = 1$ may appear. Continue to remove these new nodes until there are no nodes with a degree $k = 1$ in the network. All these removed nodes form the 1-shell, the outermost layer of the network. The degree of the remaining nodes is at least 2, and the nodes with a degree $k = 2$ are also removed until no new nodes with a degree value less than or equal to 2 appear. These removed nodes form the 2-shell. Continue to remove higher shells until all nodes are removed.

2.2.2. Generating random networks

Generate random networks according to the degree sequence of the nodes in the real network, that is, there is no change in the degree of each node. It's a permutation of edges. For each period, generate 1000 random networks and proceed with the k-shell decomposition. Every node has 1000 k-shell values generated from random networks. Then we can get a significant node whose k-shell value in the real network is higher than all 1000 k-shell values of random networks, with 0.001 as the test level.

2.2.3. Louvain algorithm

Louvain algorithm(Blondel et al., 2008) is a modularity-based approach that works well for large networks and is relatively fast. A relatively good division means that the similarity between nodes within the community is high and outside the community is low. Modularity is an indicator used to measure the quality of community division. Modularity is defined as:

$$M = (1 / 2m)\sum_{ij}\left[A_{ij} - \left(k_i k_j / 2m\right)\right]\delta\left(g_i, g_j\right) \tag{1}$$

m represents the total number of links in the network, A_{ij} is the link weight between nodes i and j, k_i and k_j are the sum of links to and from nodes i and j, $\delta(g_i, g_j)$ is equal to 1 if there is a link between nodes i and j. Otherwise, it is 0.

The Louvain algorithm can divide all nodes in the network into different communities based on modularity. In the first stage, consider each node as a single community. Calculate the modularity increment after each node and its neighbour nodes are merged into a new community. Find the largest modularity increment, and move the node to the corresponding community. Keep doing this until the maximum modularity increment is not positive. In the second stage, for each obtained community, all nodes are compressed into one node, repeating the first stage. Iterate step by step until the network reaches maximum modularity and the algorithm is stable.

2.2.4. Eigenvalue analysis

The migration matrix's eigenvalue indicates the network's ability to spread infectious diseases (Brauer et al., 2019). Once a node is deleted, the eigenvalue will inevitably decrease. Therefore, if there is a large decrease in the eigenvalue of the network after deleting a node, the deleted node is of great significance in the network in terms of disease transmission. In our work, we applied the method to every node in the network. First, calculate the eigenvalue of the whole network. Then, delete the node and its related edges, and calculate the eigenvalue of the remaining network. This way, we can get the reduction in eigenvalue from the eigenvalue of the whole network. Lastly, after repeating 366 times, we can get decreases in eigenvalue for all nodes. By comparing the decrease in eigenvalue, we can obtain the importance of each node.

2.2.5. Kendall's τ test

We applied Kendall's τ test to the k-shell value and eigenvalue decrease to find the consistency between the two parameters. Calculate the mean k-shell value and decrease of eigenvalue for each node using the above k-shell value and decrease of eigenvalue result of four periods. Compare the list of cities ranked by the mean k-shell value and the list ordered according to the mean decrease of eigenvalue with the indicator Kendall' τ. It allows a quantitative analysis of the correlations between two rankings of n objects and is given by:

$$\tau = \frac{n_c - n_d}{n(n-1)/2} \tag{2}$$

n_c is the number of pairs whose order does not change in the two different lists and n_d is the number of pairs whose order is inverted. This quantity is normalized between -1 and 1: $\tau = 1$ corresponds to an identical ranking, while $\tau = -1$ is a perfect anticorrelation, and $\tau = 0$ means the two rankings are uncorrelated.

3. Results

3.1. Nationwide k-shell values and significant nodes in four periods

Through the k-shell decomposition of the population mobility network, the k-shell value of each node is obtained, indicating the node's importance. The larger the k-shell value is, the more influential the node is for the transmission of infectious diseases in the network. Corresponding to the characteristics of the population mobility network that the east is dense and the west is sparse, the k-shell value is higher in the east region and lower in the west (Fig. 1A–D). The nodes with the highest k-shell values are listed (Table S1 Supplementary Material). As for the k-shell values for four periods, National Day has the highest nationwide mean k-shell value than the other three periods (Fig. 1E). The mean k-shell value for National Day

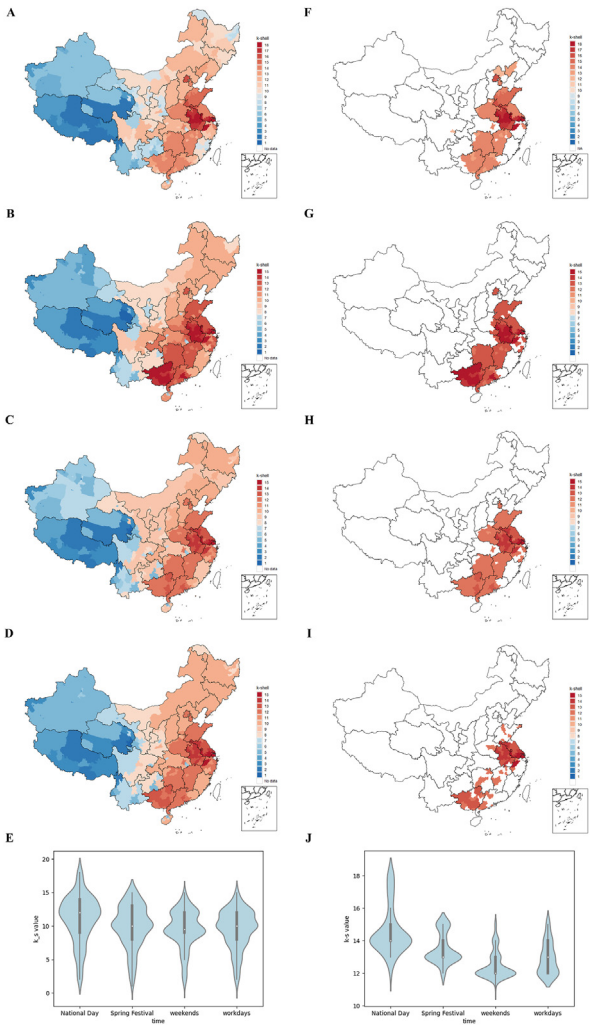

Fig. 1. K-shell values for the four periods. A-D are the nationwide k-shell values of National Day, Spring Festival, weekends, and workdays. K-shell value is higher in the east region and lower in the west. F-I are the corresponding significant nodes. The numbers of significant nodes for National Day, Spring Festival, workdays, and weekends are 122, 107, 64, and 118. The significant nodes are those with higher k-shell values. E is the violin plot for nationwide k-shell values in four periods. National Day has the highest nationwide mean k-shell value than the other three periods. J shows the k-shell values for significant nodes in four periods. The mean k-shell value of significant nodes on National Day is also the highest.

is 11.10, while the k-shell values for Spring Festival, workdays, and weekends are 9.94, 9.53, and 9.52. The difference is statistically significant with $p < 0.05$.

After generating random networks according to the degree sequence of the nodes in the real network and proceeding with k-shell decomposition, we got the significant nodes for each period. The numbers of significant nodes for National Day, Spring Festival, workdays, and weekends are 122, 107, 64, and 118 (Fig. 1F–I). The significant nodes are those with higher k-shell values. It's worth noting that Beijing is not significant on workdays and weekends (Fig. 1H and I), which is related to Beijing's strict entry-admission policy.

As shown in all four periods, city clusters are not only close in geographical location but also with almost the same k-shell values. And the k-shell values of the Yangtze River Delta, Pearl River Delta, and Beijing-Tianjin-Hebei city clusters are significantly larger than those of other regions. Thus, it's necessary to specify these clusters.

3.2. Community detection and k-shell value analysis of important regions

The results of community detection in different periods are not the same, but they are roughly divided into more or less 14 communities (Fig. 2A–D). The provinces or cities for each community are listed (Table S2 Supplementary Material). Doctor Lu Xin's group constructed a population flow network based on China's Mobile communication data(Tan et al., 2021). The

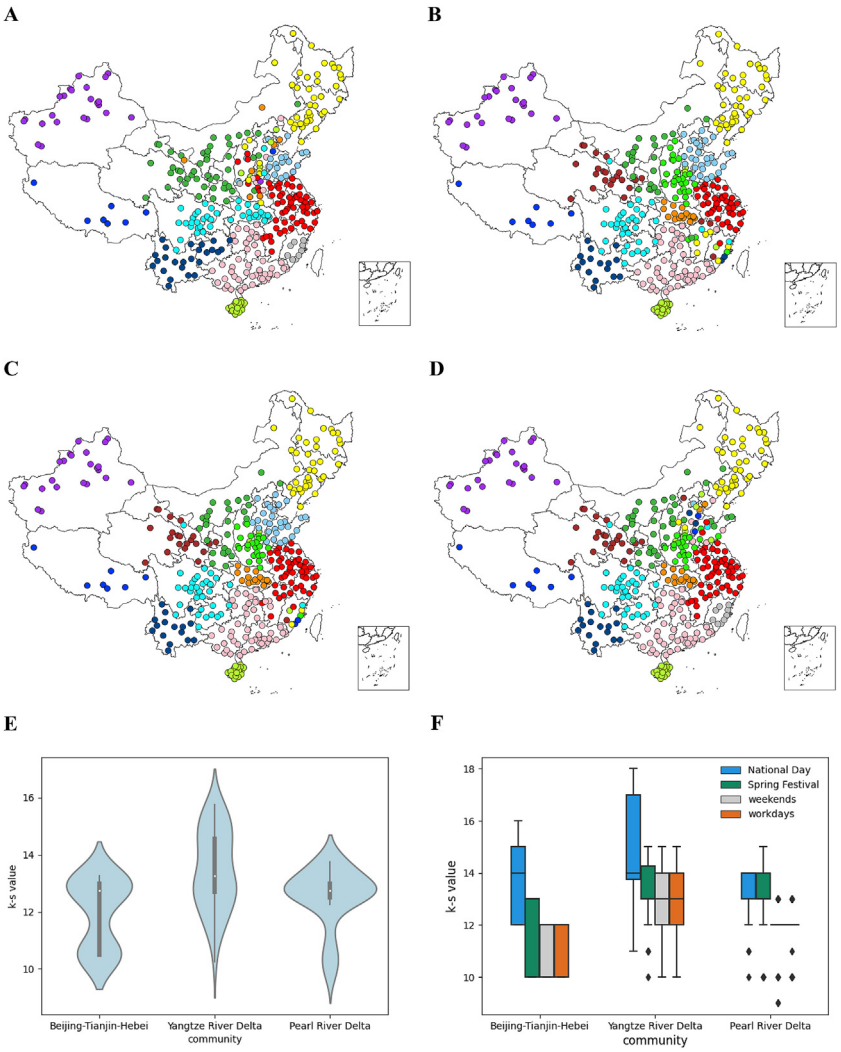

Fig. 2. Community detection for the four periods. A-D are the results of the Louvain algorithm on National Day, Spring Festival, weekends, and workdays. All 366 cities are roughly divided into more or less 14 communities. It reflects that the structure of urban agglomerations is relatively stable and does not change with the fluctuation of population flow. E shows the mean k-shell values of the three regions in four periods. Yangtze River Delta has the highest k-shell value. F is the boxplot of k-shell values for Beijing-Tianjin-Hebei, Yangtze River Delta, and Pearl River Delta in four periods. The k-shell value of the Yangtze River Delta is always higher than the other regions with $p < 0.05$.

national urban agglomeration pattern detected by the Louvain algorithm is similar to the results in our study, but the number of communities is smaller. It reflects that the structure of urban agglomerations is relatively stable and does not change with the fluctuation of population flow.

It is shown in 3.1. there exist communities centered in Beijing, Shanghai, and Guangzhou. According to the result of the Louvain algorithm, the communities are specified as Beijing-Hebei-Tianjin-Shandong (short for Beijing-Tianjin-Hebei in the following statement), Shanghai- Jiangsu-Zhejiang-Anhui-Jiangxi (Yangtze River Delta) and Guangdong-Guangxi-Hunan (Pearl River Delta), different from the official city clusters.

The k-shell values of the three communities are significantly different; for each community, the k-shell value varies with periods (Fig. 2E and F). For all periods, the k-shell value of the Yangtze River Delta is always higher than Beijing-Tianjin-Hebei and Pearl River Delta, $p < 0.05$. The k-shell value of Pearl River Delta is always higher than Beijing-Tianjin-Hebei except for National Day, $p < 0.05$. As for Beijing-Tianjin-Hebei and Yangtze River Delta, the k-shell value of National Day is the highest among all periods, $p < 0.05$, and the differences range from 1.60 to 2.55. As for Pearl River Delta, the k-shell value of National Day and Spring Festival is higher than that of workdays and weekends, and there is no significant difference between National Day and Spring Festival or workdays and weekends. The difference between National Day and weekends in Pearl River Delta is 1.63 (95%CI: 1.08–2.18).

3.3. Decrease in the eigenvalue of migration matrix on nation and community scale

The eigenvalue of the migration matrix is an applied method to represent the network's ability to transmit infectious diseases. By comparing the decrease in eigenvalue, we could obtain the importance of each node. Similar to the k-shell value distribution, the decrease in eigenvalue is larger in the eastern region and smaller in the western area (Fig. 3A–D). Cities with large decreases in eigenvalues are almost the capital of provinces like Zhengzhou, Hangzhou, Guangzhou, and so on. The twenty nodes with the largest decrease in eigenvalue for each period are listed (Table S3 Supplementary Material). Among all periods, the decrease in the eigenvalue of Shanghai maintains the largest, showing that Shanghai plays the most critical role in the transmission of infectious disease all year around. There is no significant difference between the nationwide mean decreases in four periods (Fig. 3E), $p = 0.90$. However, the decreases in eigenvalue in the Beijing-Tianjin-Hebei and Yangtze River Delta are larger than that of the Pearl River Delta (Fig. 3J), $p < 0.05$. It indicates that the other two regions have a stronger ability to transmit infectious diseases than the Pearl River Delta.

Furthermore, we analyzed the decrease in eigenvalue on the community scale to find out the core node of the community. The most influential nodes of the three communities are Jinan (Beijing-Tianjin-Hebei), Shanghai (Yangtze River Delta), and Guangzhou (Pearl River Delta). These most influential nodes remain the same in different periods (Fig. 3F–I).

3.4. Consistency between k-shell value and eigenvalue decrease

We present the list of the top 20 cities ordered according to the mean k-shell value and the mean decrease in eigenvalue in Table 1. For all 366 pairs, $\tau = 0.63$, $p < 0.001$. The result indicates a significant correlation between the mean k-shell value and the mean decrease in eigenvalue. It can be concluded from such consistency that the nodes with higher k-shell values have a greater impact on infectious disease transmission since the eigenvalue decrease represents the node's impact on the network.

3.5. Association with infectious disease transmission

As the k-s value and eigenvalue decrease of a city reflect its influence on the network, it may

influence a wider range with a higher k-s value in the aspect of infectious disease transmission. Therefore, we analyzed the correlation between the spread range of a city and its k-s value as well as eigenvalue decrease to prove the value of these methods in infectious disease transmission.

Within 14 days, Jilin, Lanzhou, Shanghai, and Quanzhou could affect 7, 4, 11, and 8 cities respectively (Fig. 4A). We found that both k-s value and eigenvalue decrease were highly correlated with number of affected cities (Fig. 4B). K-s value shows a significant association with number of affected cities ($r = 0.98$, $p = 0.02$), while eigenvalue decrease has a weaker association with number of affected cities ($r = 0.90$, $p = 0.09$). In other words, k-s value has a better prediction of the spread range of infectious diseases. Combining population epidemic model(Fan et al., 2020), we analyzed the Spearman correlation between risk rankings and k-shell values as well as eigenvalue decrease (Fig. S1 Supplementary Material). It shows that cities with higher k-shell values and larger decreases of maximum eigenvalue have a higher risk, and the correlation between risk rankings and k-shell values is 0.57 ($p < 0.001$), while the correlation between risk rankings and maximum eigenvalue decrease is 0.65 ($p < 0.001$). It indicates that eigenvalue decrease performs slightly better in risk assessment. It may be appropriate to integrate k-shell value and eigenvalue decrease to make better predictions about infectious disease transmission.

4. Discussion

In our current work, we analyzed the spatiotemporal features of the population mobility network through k-shell decomposition, random network, Louvain algorithm, eigenvalue analysis, and Kendall's τ test. K-shell value is the highest on

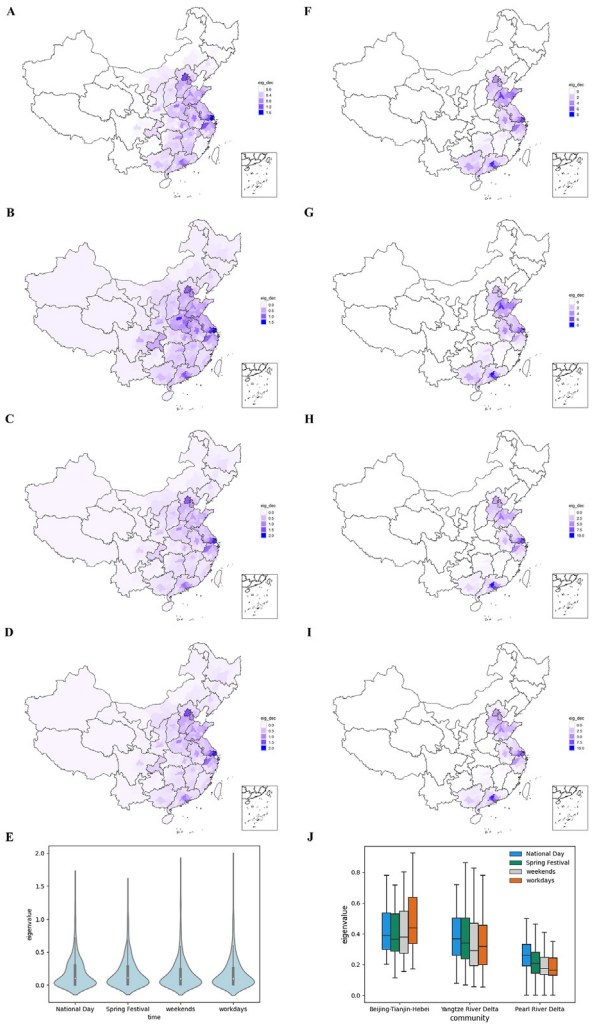

Fig. 3. Decrease in eigenvalue in four periods. A-D are the nationwide decrease in the eigenvalue on National Day, Spring Festival, weekends, and workdays. Similar to the k-shell value distribution, the decrease in eigenvalue is larger in the eastern region and smaller in the western area. Shanghai maintains the largest decrease in eigenvalue. F-I are the corresponding decrease of eigenvalue on the scale of community. The most influential nodes for the three communities are Jinan, Shanghai, and Guangzhou. These nodes remain the same in different periods. E shows no significant difference in the nationwide mean decreases between four periods. J is the boxplot of the decrease in eigenvalue in the three regions in four periods. The decreases in eigenvalue in Beijing-Tianjin-Hebei and Yangtze River Delta are significantly larger than those of the Pearl River Delta.

National Day and in the Yangtze River Delta among all periods and communities. According to the decrease in eigenvalue, there is a higher risk of outbreak and infectious disease transmission on National Day and Spring Festival on weekends and workdays. And Yangtze River Delta is at the highest risk of an outbreak all year around. Among all 366 cities, Shanghai has the highest k-shell value as well as eigenvalue decrease in all periods. We also have found the consistency between k-shell value and eigenvalue decrease.

Our work is practical in control strategy making and epidemic prevention. Nodes with higher k-shell values need extra attention, like Shanghai and Guangzhou. Once there is an outbreak of infectious disease in the core node, the transmission may be too fast to take a timely response. Similarly, it's efficient to bring core nodes under control to reduce transmission and infected nodes. It's demonstrated through numerical simulation and ER model networks that immunizing core nodes in the network can effectively reduce the number of infected nodes and sufficiently reduce the network density to control the outbreak of infectious diseases(Saxena et al., 2018). We also work out city clusters through the Louvain algorithm to illustrate mobility patterns in different periods. All 366 cities are roughly divided into fourteen communities, more than the number of communities detected by Doctor Lu Xin's group(Tan et al., 2021). But the national agglomeration pattern is similar. According to the principle of the Louvain algorithm, nodes in the same community are more strongly attached than nodes in other communities, facing a higher risk of infection once there is an outbreak inside the community. It can be employed to create an effective and accurate health surveillance system that grows as time passes(Elgazzar et al., 2021).

Table 1
The list of top 20 cities is ordered according to the mean k-shell value and mean decrease of maximum eigenvalue.

Ranking	K-s value	Decrease of eigenvalue
1	Shanghai	Shanghai
2	Nanjing	Suzhou
3	Suzhou	Beijing
4	Hangzhou	Hangzhou
5	Hefei	Shenzhen
6	Wuhu	Guangzhou
7	Chuzhou	Zhengzhou
8	Fuyang	Dongguan
9	Lu'an	Nanjing
10	Wuxi	Foshan
11	Chizhou	Hefei
12	Bengbu	Wuxi
13	Maanshan	Jinan
14	Changzhou	Tianjin
15	Tongling	Huizhou
16	Anqing	Qingdao
17	Xuzhou	Langfang
18	Huainan	Ningbo
19	Suzhou	Jiaxing
20	Bozhou	Wuhan

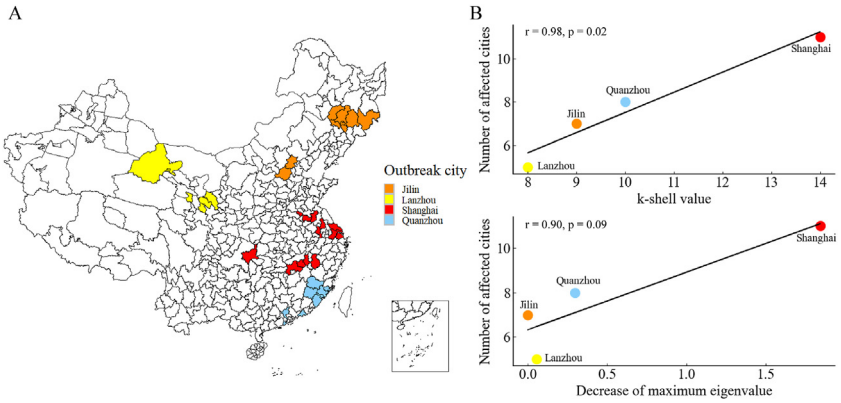

Fig. 4. Correlation between number of affected cities and k-s value, decrease of maximum eigenvalue. A shows four origin cities and their corresponding affected cities. Jilin, Lanzhou, Shanghai, and Quanzhou could affect 7, 4, 11, and 8 cities respectively within 14 days. B shows the Pearson correlation between the number of affected cities and k-shell value, as well as eigenvalue decrease. K-shell value shows a significant association with the number of affected cities ($r = 0.98$, $p = 0.02$).

Complex network analysis is challenging due to massive data, network size, and changing topology. Although k-shell value performs better in identifying critical nodes than other metrics like betweenness centrality and closeness centrality, it does have backdraws in real network analysis. K-shell value always assigns many nodes with the same value, leading to the limitation in distinguishing the influences of these nodes(Li & Huang, 2021). In our work, 366 cities are assigned into no more than twenty k-shell layers. Many nodes have the same k-shell values, especially those with higher k-shell values, making it hard to portray significance exactly. And it's rarely accurate in quantifying the spreading power of nodes that are not highly influential(Lawyer, 2015). By the way, depending on a single characteristic of nodes to reliably identify influential spreaders is inadequate(Li & Huang, 2022; Ullah et al., 2021). Thus, k-shell value should be adopted to form a comprehensive indicator combined with other metrics like degree centrality, betweenness centrality, gravitational centrality(Wang et al., 2018), and so on. As for Louvain algorithm, it may yield arbitrarily badly connected communities(Traag et al., 2019). Although in our work, there exists no such problem, it's worth noting when applying Louvain algorithm to other networks.

Our analysis finds out significant cities and communities in different periods. According to k-shell value and eigenvalue decrease, there is a higher risk of an outbreak on National Day and Spring Festival. Shanghai and Yangzte River Delta are core regions in infectious disease transmission. We also discover the consistency between k-shell value and eigenvalue decrease. Modelling disease transmission needs to take this complex spatiotemporal information into account to make more accurate predictions about the transmission speed, arrival time, and so on. Further modelling and analytical work will be conducted to refine and generalize this information as part of the model.

CRediT authorship contribution statement

Wenjie Li: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ye Yao:** Conceptualization, Formal analysis, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Shanghai Municipal Health Commission Clinical Research Program (20214Y0020), the General Program of Natural Science Foundation of Shanghai Municipality (22ZR1414600), and the Young Health Talents Program of Shanghai Municipality (2022YQ076).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2023.10.003.

References

de Arruda, G. F., Barbieri, A. L., Rodríguez, P. M., Rodrigues, F. A., Moreno, Y., & Costa Lda, F. (2014). Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics, 90*(3), Article 032812. https://doi.org/10.1103/PhysRevE.90.032812

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), Article P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). In F. Brauer, C. Castillo-Chavez, & Z. Feng (Eds.), *Mathematical models in epidemiology* (1 ed.). Springer. https://doi.org/10.1007/978-1-4939-9828-9.

Buchel, O., Ninkov, A., Cathel, D., Bar-Yam, Y., & Hedayatifar, L. (2021). Strategizing COVID-19 lockdowns using mobility patterns. *Royal Society Open Science, 8*(12), Article 210865. https://doi.org/10.1098/rsos.210865

Bucur, D., & Holme, P. (2020). Beyond ranking nodes: Predicting epidemic outbreak sizes by network centralities. *PLoS Computational Biology, 16*(7), Article e1008052. https://doi.org/10.1371/journal.pcbi.1008052

Chaharborj, S. S., Nabi, K. N., Feng, K. L., Chaharborj, S. S., & Phang, P. S. (2022). Controlling COVID-19 transmission with isolation of influential nodes. *Chaos, Solitons & Fractals, 159*, Article 112035. https://doi.org/10.1016/j.chaos.2022.112035

Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in social networks: Using network analysis to identify high-risk individuals. *American Journal of Epidemiology, 162*(10), 1024–1031. https://doi.org/10.1093/aje/kwi308

Clusella, P., Grassberger, P., Pérez-Reche, F. J., & Politi, A. (2016). Immunization and targeted destruction of networks using explosive percolation. *Physical Review Letters, 117*(20), Article 208301. https://doi.org/10.1103/PhysRevLett.117.208301

Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the U S A, 103*(7), 2015–2020. https://doi.org/10.1073/pnas.0510525103

Dye, C., & Gay, N. (2003). Epidemiology. Modeling the SARS epidemic. *Science, 300*(5627), 1884–1885. https://doi.org/10.1126/science.1086925

Elgazzar, H., Spurlock, K., & Bogart, T. (2021). Evolutionary clustering and community detection algorithms for social media health surveillance. *Mach Learn Appl, 6*, Article 100084. https://doi.org/10.1016/j.mlwa.2021.100084

Fan, C., Cai, T., Gai, Z., & Wu, Y. (2020). The relationship between the migrant population's migration network and the risk of COVID-19 transmission in China—empirical analysis and prediction in prefecture-level cities. *International Journal of Environmental Research and Public Health, 17*(8).

Freeman, L. C. (1977). Set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41. https://doi.org/10.2307/3033543

Henao-Restrepo, A. M., Preziosi, M. P., Wood, D., Moorthy, V., & Kieny, M. P. (2016). On a path to accelerate access to Ebola vaccines: The WHO's research and development efforts during the 2014-2016 Ebola epidemic in West Africa. *Current Opinion in Virology, 17*, 138–144. https://doi.org/10.1016/j.coviro.2016.03.008

Joshi, Y. V., & Musalem, A. (2021). Lockdowns lose one third of their impact on mobility in a month. *Scientific Reports, 11*(1), Article 22658. https://doi.org/10.1038/s41598-021-02133-1

Khan, K., Arino, J., Hu, W., Raposo, P., Sears, J., Calderon, F., Heidebrecht, C., Macdonald, M., Liauw, J., Chan, A., & Gardam, M. (2009). Spread of a novel influenza A (H1N1) virus via global airline transportation. *New England Journal of Medicine, 361*(2), 212–214. https://doi.org/10.1056/NEJMc0904559

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics, 6*(11), 888–893. https://doi.org/10.1038/nphys1746

Krackhardt, D. (1990). Assessing the political landscape - structure, cognition, and power in organizations. *Administrative Science Quarterly, 35*(2), 342–369. https://doi.org/10.2307/2393394

Lawyer, G. (2015). Understanding the influence of all nodes in a network. *Scientific Reports, 5*, 8665. https://doi.org/10.1038/srep08665

Li, Z., & Huang, X. (2021). Identifying influential spreaders in complex networks by an improved gravity model. *Scientific Reports, 11*(1), Article 22194. https://doi.org/10.1038/s41598-021-01218-1

Li, Z., & Huang, X. (2022). Identifying influential spreaders by gravity model considering multi-characteristics of nodes. *Scientific Reports, 12*(1), 9879. https://doi.org/10.1038/s41598-022-14005-3

Louie, J. K., Acosta, M., Jamieson, D. J., & Honein, M. A. (2010). Severe 2009 H1N1 influenza in pregnant and postpartum women in California. *New England Journal of Medicine, 362*(1), 27–35. https://doi.org/10.1056/NEJMoa0910444

Mukherjee, P., Lim, P. L., Chow, A., Barkham, T., Seow, E., Win, M. K., Chua, A., Leo, Y. S., & Cheng Chen, M. I. (2010). Epidemiology of travel-associated pandemic (H1N1) 2009 infection in 116 patients, Singapore. *Emerging Infectious Diseases, 16*(1), 21–26. https://doi.org/10.3201/eid1601.091376

Müller, M. A., Meyer, B., Corman, V. M., Al-Masri, M., Turkestani, A., Ritz, D., Sieberg, A., Aldabbagh, S., Bosch, B. J., Lattwein, E., Alhakeem, R. F., Assiri, A. M., Albarrak, A. M., Al-Shangiti, A. M., Al-Tawfiq, J. A., Wikramaratna, P., Alrabeeah, A. A., Drosten, C., & Memish, Z. A. (2015). Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: A nationwide, cross-sectional, serological study. *The Lancet Infectious Diseases, 15*(5), 559–564. https://doi.org/10.1016/s1473-3099(15)70090-3

Saxena, C., Doja, M. N., & Ahmad, T. (2018). Group based centrality for immunization of complex networks. *Physica A: Statistical Mechanics and Its Applications, 508*, 35–47. https://doi.org/10.1016/j.physa.2018.05.107

Shaw, K. (2006). The 2003 SARS outbreak and its impact on infection control practices. *Public Health, 120*(1), 8–14. https://doi.org/10.1016/j.puhe.2005.10.002

Tan, S., Lai, S., Fang, F., Cao, Z., Sai, B., Song, B., Dai, B., Guo, S., Liu, C., Cai, M., Wang, T., Wang, M., Li, J., Chen, S., Qin, S., Floyd, J. R., Cao, Z., Tan, J., Sun, X., … Lu, X. (2021). Mobility in China, 2020: A tale of four phases. *National Science Review, 8*(11), nwab148. https://doi.org/10.1093/nsr/nwab148

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports, 9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z

Ullah, A., Wang, B., Sheng, J., Long, J., Khan, N., & Sun, Z. (2021). Identification of nodes influence based on global structure model in complex networks. *Scientific Reports, 11*(1), 6173. https://doi.org/10.1038/s41598-021-84684-x

Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine and International Health, 25*(3), 278–280. https://doi.org/10.1111/tmi.13383

Wang, J., Li, C., & Xia, C. (2018). Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Applied Mathematics and Computation, 334*, 388–400. https://doi.org/10.1016/j.amc.2018.04.028

Wei, X., Zhao, J., Liu, S., & Wang, Y. (2022). Identifying influential spreaders in complex networks for disease spread and control. *Scientific Reports, 12*(1), 5550. https://doi.org/10.1038/s41598-022-09341-3

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. *Lancet, 395*(10225), 689–697. https://doi.org/10.1016/s0140-6736(20)30260-9

Zhang, S. X., Chen, J., Afshar Jahanshahi, A., Alvarez-Risco, A., Dai, H., Li, J., & Patty-Tito, R. M. (2022). Succumbing to the COVID-19 pandemic-healthcare workers not satisfied and intend to leave their jobs. *International Journal of Mental Health and Addiction, 20*(2), 956–965. https://doi.org/10.1007/s11469-020-00418-6

Zhao, D., Yang, S., Han, X., Zhang, S., & Wang, Z. (2020). Dismantling and vertex cover of network through message passing. *IEEE Transactions on Circuits and Systems II: Express Briefs, 67*(11), 2732–2736. https://doi.org/10.1109/TCSII.2020.2973414

Zhu, Z., & Zhang, Y. (2019). Factors affecting the spread of multiple information in social networks. *PLoS One, 14*(12), Article e0225751. https://doi.org/10.1371/journal.pone.0225751