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

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

Mingtao Li a,*, Jin Cui a, Juan Zhang b, Xin Pei a, Guiquan Sun b,c

Article history:
Received 14 July 2022
Received in revised form 30 September 2022
Available online 14 October 2022

Keywords:
COVID-19
Contact network
Centrality
Link prediction
Simulated propagation

1. Introduction

The outbreak of 2019 novel coronavirus pneumonia (COVID-19) caused by the coronavirus SARS-CoV-2 has had a profound impact on people's lives around the world. Since the end of 2019, the spread of COVID-19 has brought enormous challenges to public health. How to effectively and quickly control the epidemic has become an urgent problem to be solved. The launch of multiple vaccines has brought hope to people's lives [1]. However, while vaccines are being developed, viruses are constantly evolving and mutating [2], which exposes the world to potentially huge risks.

Since the outbreak of the novel coronavirus pneumonia, China has adopted various non-drug interventions (closure of cities, villages, tracking and quarantine, etc.) to control the spread of the epidemic. The implementation of various measures made the epidemic in other parts of China, except Hubei Province, basically controlled in mid-to-early March 2020. Since the travel trajectory [3] of each confirmed case will be announced in the early stage of the epidemic in China, the contacts of the case can be identified based on its travel trajectory. An almost clear cases’ contact network can be obtained, and which provides a more realistic network for studying the spread of COVID-19.

Studying the effectiveness of its early prevention and control strategies can provide more reasonable suggestions for China’s current epidemic prevention and control. There have been many studies based on the mathematical model to
study the effectiveness of prevention and control strategies for COVID-19 [4–9]. Tang et al. [6] devised SEIR model on the estimation of the transmission risk of COVID-19 and showed the effectiveness of control strategy by intensive contact tracing followed by quarantine and isolation in mainland China. Li et al. [7] developed a SEIQR difference-equation model of COVID-19 in Shanxi province that takes into account the transmission with discrete time imported cases. Sun et al. [8] presented a dynamical model to show the propagation of COVID-19 in Wuhan and the effects of lock-down and medical resources. Li et al. [9] developed differential equation model with tracing isolation strategy with close contacts of newly confirmed cases and discrete time imported cases to make predictions and perform assessment and risk analysis for COVID-19 outbreaks in Tianjin and Chongqing city. In addition, many scholars have also done research on individual-based network dynamics models. Xu et al. [10] established an individual-based transmission model to simulate the impact of different levels of non-drug interventions on controlling the spread of COVID-19. The results showed that isolating the infected and their first-level close contacts could not effectively control the spread of the disease. Isolating the second-level close contacts can reduce the size of the outbreak. Firth et al. [11] simulated the prevention and control strategy of COVID-19 in a real social network, and the results showed that tracing of contacts reduced the outbreak scale of the disease, and maintaining social distancing while adopting the tracing and isolation strategy was very effective for epidemic control. Therefore, understanding and studying detailed cases’ transmission contact network can show the main factors influencing COVID-19 outbreaks.

Using information published by the Health Commission of Tianjin city [12], we firstly constructed a case contact network data-set based on the statistical 136 confirmed COVID-19 patients. The detailed information includes the cumulative and daily laboratory-confirmed cases, and life track of these laboratory-confirmed cases. Then the topological features of this contact network were analyzed, and further we used link prediction algorithm to improve the original case contact network. Finally, we established a susceptible–latent–exposed–infected–recover (SLEIR) transmission model based on individual transmission to study the influence of the central nodes on COVID-19 transmission.

2. The data and network

2.1. The data

The first COVID-19 confirmed case appeared in Tianjin on January 21, 2020 [12]. Once a confirmed case is found, epidemiological investigators will investigate their movement track and close contacts, and then provide scientific basis for relevant departments and governments to prevent and control the epidemic through statistical analysis of convection data. The cumulative number of confirmed cases in Tianjin had reached 136 by March 14, 2020. According to the diagnosis time and residence area of 136 confirmed cases, the chart of daily newly confirmed cases and the map of confirmed cases in Tianjin are shown in Fig. 1. The time series chart of cases in Tianjin (Fig. 1(a)) showed that the daily newly confirmed cases fluctuated greatly after January 31, 2020, which could be caused by the department store incident in Baodi District. Ignoring the 6 confirmed cases from outside Tianjin, we found that Baodi District had the most cumulative confirmed cases, accounting for 46% of Tianjin, followed by Hedong District and Hebei District, accounting for 12% and 9% of Tianjin, respectively. Fig. 1(b) showed that Baodi District has the most cumulative number of confirmed cases in the early stage of Tianjin, followed by Hedong District and Hebei District, which is consistent with our calculation.

2.2. The network

From the Municipal Health Commission of Tianjin [12], we obtain detailed information of confirmed cases, such as gender, age, place of residence, date of onset, date of diagnosis, track of travel and history of residence or travel in Wuhan. Since the early stage of COVID-19 is characterized by contact transmission, and the existing data contains the contact information of confirmed cases, the contact data of cases can be used to construct a case contact network to study the spread of COVID-19 on the network. However, the transmission direction of the disease cannot be determined simply based on the date of diagnosis, because the incubation period of the disease varies from person to person. For example, the average incubation period of the disease obtained in literature [13,14] are 5.2 days and 4 days, respectively. Therefore, the network studied in this paper is an undirected network.

Then we build the case contact network. First, case nodes were numbered $s_1, s_2, \cdots, s_{136}$ according to its confirmed sequence. Secondly, we used their contact information to build links between nodes. If there is contact between case 1 and case 2, there exits a link between them. By sorting out the existing data we can obtain a network with 148 links is shown in Fig. 2 (there are 15 isolated nodes in the constructed network). Table 1 shows the partial properties of this constructed network by using method of [15]. The values of some properties show that this network is a sparse network. Due to the lack of information, there are many isolated nodes in the network. Therefore, the network properties reflected by average degree, average path length and clustering coefficient are not discussed here. However, the network diameter is much larger than the network average path length, which means that when the disease spreads in the network, the number of people that can be spread through the diameter is more than twice the number of people spread through the average path.
Fig. 1. Chart of daily newly confirmed cases and the map of confirmed cases in Tianjin. (a) New daily cases. (b) The map of confirmed cases. The darker the color, the more cases.
2.3. The degree and cumulative degree distribution of the network

In order to understand the characteristics of the case contact network, we then studied the degree distribution of the network. Degree distribution refers to the proportion $p_k$ of nodes with degree $k$ in the number of nodes in the whole network. The degree distribution diagram of the network is shown in Fig. 3(a).

From Fig. 3(a), it can be seen that the number of nodes in the case contact network with degree 1 is the most, followed by the two-degree node, the three-degree node and the zero-degree node respectively. The case contact network is characterized by the majority of small degree nodes and the few large degree nodes. Ignoring nodes with degree 0, degree distribution roughly follows long-tail distribution, as shown in Fig. 3(a). Because the degree distribution of the network is not regular, the cumulative degree distribution is introduced to analyze the network. The cumulative degree distribution is expressed by $P(k)$, which refers to the probability that the node degree is greater than or equal to $k$. The cumulative degree distribution corresponding to the case contact network is shown in Fig. 3(b), which shows that the cumulative degree distribution of the case contact network roughly follows a power-law distribution, $k^{-\gamma}$, where $\gamma$ is called the power index [16], and is equal to 1.403538. The cumulative degree distribution values first decreased rapidly, which was caused by many small degree nodes in the case contact network. The results indicated that the node (s37) with 28-degree is dominant in the case contact network, whose link number (28) accounts for 19% of the total link number (148).

3. Network centrality index and centrality analysis

In the established case contact network, it is necessary to introduce a centrality index to analyze the network in order to find the nodes that have the greatest influence on the network. Centrality measures are used to measure the importance of nodes in a network. In this study, degree centrality, betweenness centrality, closeness centrality and eigenvector centrality were used to measure the importance of nodes, whose calculation formula are shown in Table 2.
Fig. 3. The degree distribution and cumulative degree distribution of the case contact network. $k$: the degree of node, $p(k)$: the proportion of nodes with a degree of $k$ to the number of nodes in the entire network, $P(k)$: the proportion of nodes whose degree is greater than or equal to $k$ to the number of nodes in the entire network. (a) Degree distribution. (b) The cumulative degree distribution.

Table 2
Centrality measures.

| Centrality measures          | Formula                                      |
|------------------------------|----------------------------------------------|
| Degree centrality [17,18]    | $C_D(i) = \sum_{j=1}^{N} a_{ij}$            |
| Betweenness centrality [18]  | $C_B(i) = \sum_{s \neq i \neq t} \frac{n_{st}}{d_{st}}$ |
| Closeness centrality [19]    | $C_C(i) = \frac{1}{\sum_{j \neq i} d_{ij}}$ |
| Eigenvector centrality [20]  | $C_E(i) = c \sum_{j=1}^{N} a_{ij} x_j$       |

Table 3
Top ten nodes by degree, betweenness, closeness, eigenvector.

| Rank | $C_D$ | $C_B$ | $C_C$ | $C_E$ | Rank | $C_D$ | $C_B$ | $C_C$ | $C_E$ |
|------|-------|-------|-------|-------|------|-------|-------|-------|-------|
| 1    | s37   | s37   | s37   | s37   | 6    | s5    | s127  | s119  | s124  |
| 2    | s6    | s34   | s34   | s43   | 7    | s103  | s43   | s85   | s85   |
| 3    | s43   | s20   | s43   | s34   | 8    | s119  | s5    | s122  | s99   |
| 4    | s34   | s6    | s20   | s48   | 9    | s2    | s83   | s48   | s53   |
| 5    | s93   | s119  | s87   | s128  | 10   | s8    | s85   | s51   | s51   |

By calculation, Table 3 lists the top 10 nodes for four types of centrality. It can be found that the first node of four kinds of centrality is the same node (s37), which indicates that the role of the case cannot be ignored in the case contact network. In addition, we also found that nodes s34 and s43 frequently appeared among the top ten nodes of the four centrality rankings, which was consistent with the existing data, and were all related to the department store of Baodi District. That is to say, s34, s37 and s43 played an important role in the epidemic in Baodi District. They were all sales staff of the department store in Baodi District, and greatly affected the spread of the COVID-19 in Tianjin. The top three nodes of centrality are presented in the network diagram as shown in Fig. 4.

The regions of the top 10 nodes of the four kinds of centrality are analyzed below. Fig. 5 shows the geographical distribution of the top 10 nodes of centrality in Tianjin. It is not difficult to find that most of the top ten nodes belong to Baodi District, which is consistent with the fact that Baodi District is the largest cluster of COVID-19 epidemic in Tianjin. Both Hedong District and Hebei District appear in Fig. 5(a) and Fig. 5(b). This suggests that cases of Hedong district and Hebei district of Tianjin COVID-19 outbreak are influential in the development process. This was caused by the work area of the first confirmed case in Hedong District, who worked in the Tianjin bullet train section, traveled to Wuhan on business, and then infected colleagues living in Hedong District and Hebei District. This is consistent with the statistical data of Tianjin, namely, Baodi District of Tianjin has the most cases, followed by Hedong district and Hebei District (Fig. 1(b)).

4. Link prediction

4.1. Introduction to link prediction

Because the established case contact network may contain false links and missing links. Therefore, link prediction algorithms are needed to improve the network. The basic idea of link prediction [21] is to assign a fraction $s_{xy}$ to each pair
of unlinked nodes \((x, y)\) in an undirected network \(G(V, E)\) with \(N\) nodes and \(M\) links, and then order all unconnected node pairs from the largest to the smallest according to this value, and the node pairs at the top have the highest probability of appearing links.

In order to test the accuracy of link prediction algorithm, the known links in the network set \(E\) is usually divided into training set \(E^T\) and testing set \(E^P\): \(E = E^T \cup E^P, E^T \cap E^P = \emptyset\). Only the information of the testing set is used in the calculation, and any possible link between any pair of nodes that do not belong to the existing link set \(E\) is called a non-existent link. Two commonly used metrics to measure the accuracy of link prediction algorithms are AUC and PRECISION. AUC is introduced here.

AUC measures the accuracy of the algorithm as a whole, can be understood as the probability that the score of a link in the testing set is higher than the score of a randomly selected link that does not exist [21]. In other words, randomly choose an edge \((L1)\) from the test set to compare with a non-existent edge \((L2)\): if the score value of \(L1\) is greater than the score value of \(L2\), then add 1 point; if the scores are equal, add 0.5 points. So independently compare \(n\) times, if there is \(n'\) times the score value of \(L1\) is greater than \(L2\), \(n''\) times is equal, then AUC is defined as:

\[
AUC = \frac{n' + 0.5n''}{n}.
\]  

(1)

The larger the AUC value is, the more accurate the prediction result is.

Since few case attributes can be obtained in the case contact network, node similarity index based on local information is adopted here to make link prediction for the existing network. CN index, AA index and RA index are adopted in this paper. The specific formula is shown in Table 4.
Fig. 5. The geographical distribution of the top ten nodes of centrality in each district of Tianjin. The darker the color of the region, the greater the proportion of the top ten nodes belonging to that region. (a) Degree centrality. (b) Betweenness centrality. (c) Closeness centrality. (d) Eigenvector centrality.

Table 4

| Index | Formula |
|-------|---------|
| CN [22] | \(s_{xy} = I(x) \cap I(y)\) |
| AA [23] | \(s_{xy} = \sum_{z \in I(x) \cap I(y)} \frac{1}{k_z}\) |
| RA [24] | \(s_{xy} = \sum_{z \in I(x) \cap I(y)} \frac{1}{k_z}\) |

4.2. Data processing of case contact network

Since there are many isolated points in the constructed case contact network, in order to solve this problem, the following assumptions are made:

(1) Linking the edges according to the region where the isolated nodes are located, that is, it is assumed that they have contacts with all confirmed cases in the same region;
Fig. 6. Optimized case contact network, which has 136 nodes and 197 links. Red nodes are isolated nodes in the network shown in Fig. 3.

Table 5

| Index | CN   | AA   | RA   |
|-------|------|------|------|
| AUC   | 0.8519 | 0.8608 | 0.8609 |

(2) When there is only one case in the region where the isolated node is located, the confirmed cases in the region closest to the region are preferred to have contacts with the isolated node.

Under the above assumptions, a case contact network with 136 nodes and 197 edges was constructed, as shown in Fig. 6.

4.3. Reconstructed network after link prediction

The optimized case contact network is used to make link prediction. Here, it is assumed that the ratio of training set to testing set is 9 : 1. The reconstructed network after link prediction using the Common Neighbor (CN) index, Adamic–Adar (AA) index and Resource Allocation (RA) index are shown in Fig. 7.

After observing the reconstructed network predicted by the three indexes, it can be found that all they contain two isolated nodes (s15, s30). Compared with the existing data, since these two isolated nodes are imported cases, this paper does not analyze whether these two cases have contact with other cases. The AUC values are used to analyze the prediction accuracy of the three indicators. After 10000 simulations, the mean AUC values of the three indicators are obtained, as shown in Table 5.

It can be concluded that the prediction accuracy of RA index is the highest. Therefore, the predicted results of RA indicators were compared with the original case contact network. Through the link prediction, the number of links in the network changes from 197 to 584. It is speculated that the extra links are mostly contacts of family members and colleagues that have not been counted. And the properties of the reconstructed network are shown in Table 6. The reconstructed network has a small average path length and large clustering coefficient, so the network after link prediction has small-world effect [25,26], which indicates that the spread of COVID–19 is highly aggregative. By link prediction, the network diameter decreased significantly from 10 to 9 (s106 – s3 – s9 – s19 – s63 – s29 – s20 – s119 – s41 – s110). This means that when the number of network links is increased by link prediction, the transmission chain may not become shorter, which may be related to the incident of Baodi District department store, i.e. infection of staff, customer infection, customer family infection, etc.
Fig. 7. Reconstructed network. The thickness of the network edge is different because the weight of the edge is different for different link prediction indexes. The larger the circular node indicates the greater the degree of the node. Red square nodes are isolated nodes. (a) Common Neighbor (CN). (b) Adamic-Adar (AA). (c) Resource Allocation (RA).

Table 6

| Network properties (link prediction results of RA index) | Value   |
|--------------------------------------------------------|---------|
| Density                                                | 0.06361656 |
| Average degree                                         | 8       |
| Average path length                                    | 3.262306  |
| Clustering coefficient                                 | 0.8294807 |
| Diameter                                               | 9       |

5. Network dynamics analysis based on individual model

In order to study the influence of the central nodes on COVID-19 transmission in the network, a susceptible–latent–exposed–infected–recover (SLEIR) transmission model based on individual transmission was established based on probabilistic discrete-time Markov chains [27,28].
equations with Markov chain are given by the following.

where

Hence, the transmission from exposed individual and infected individual can be obtained as follows:

where $a_{ij}$ indicates the link from $i$ to $j$ in the network, and $R = (a_{ij})_{N \times N}$ is the adjacency matrix of the network.

Then, we can use the following equation to calculate the probability $\lambda(j, t)$ of a susceptible individual $j$ being not infected by its exposed or infected individual neighbors at time $t$.

Furthermore, the probability $\xi(j, t)$ of a susceptible individual $j$ being infected by its exposed or infected individual neighbors at time $t$ can be obtained:

Hence, the transmission rate is $\xi(j, t) P_{ik}^I(t)$.

5.2. Impact of central nodes on disease transmission

According to the existing data consider the influence of central nodes on the spread of disease in the network, it is assumed that the initial infected nodes are $s_1, s_3, s_4$ and the first three nodes of the central nodes (degree: $s_37, s_43, s_6$; betweenness: $s_37, s_34, s_20$; closeness: $s_37, s_34, s_43$; eigenvector: $s_37, s_43, s_34$). Without loss of generality, disease transmission is simulated on the original network and the predicted network, respectively. As the contact transmission rate of the disease is unknown, it is assumed that $\beta_{E} = \beta_{I} = 0.4998$, and the transfer rate $\delta, \sigma, \gamma$ are $1/2.5, 1/3, 1/4$ [9]. Through simulations, we obtained the mean of the peak scale size of the disease as shown in Table 7.

As can be seen from Table 7, in both the original network and the prediction network, the disease scale caused by the betweenness nodes among the central nodes is the largest, indicating that the media has a great influence in the process of network transmission. In addition, we also found that the peak scale of disease caused by the initial infected nodes in the original network and the prediction network was significantly different. In order to clarify the cause of this disparity,
we calculated the shortest path from the prediction network to the central node, and found that after link prediction, node s1 could reach the central node through the shortest path (s1 − s24 − s49 − s37 − s34, s1 − s24 − s49 − s37 − s43, s1 − s24 − s49 − s29 − s20 − s22 − s6), thus causing the peak scale of the disease to be larger than that caused by the central node. From the peak time of the disease, the four central nodes no difference. However, compared with the peak propagation of the original nodes in the two networks, it can be found that the nodes can reach the peak faster when there are fewer connected paths. This means that when a disease is spreading through the network, the necessary disconnection helps the disease peak quickly, which in turn reduces its spread. The change curve of simulated infected persons is shown in Fig. 8.

Fig. 8(a) shows that the first three nodes (s37, s34, s20) of betweenness centrality is more likely to cause the large-scale spread of the disease. When the initial infected node in the network is the actual node data(s1,s3,s4), the peak size of the disease is much smaller than that caused by the central node, which means that the timely discovery and find out the key nodes in the network has a great influence on the spread of disease, and to find the real key nodes in the spread of the disease, usually depends on the nature of work and the social network. Fig. 8(b) shows that the propagation scale caused by the initial node is basically the same as that caused by the central node. This is because link prediction changes the link of s1 in the initial node, and then it can reach the central node through some nodes, thus driving the propagation of the central node. This further reflects the influence of central nodes on the spread of disease in the network. That is, for the spread of disease in the network, it is necessary to find the central nodes with high aggregation and cut off the path to the central nodes in time, so that the disease can be controlled in a short time.

5.3. Impact of transmission rate on disease transmission

Due to the various prevention and control strategies adopted during the spread of COVID-19 to reduce its transmission rate of contact, such as 14-day isolation of infected persons, wearing masks and maintaining social distancing. Therefore, the influence of infection rate on disease transmission was further considered. To facilitate comparison, we chose to simulate the impact of transmission rate on disease transmission in the prediction network. Due to the large scale of diseases caused by initial nodes and media-central nodes, we considered the influence of transmission rate on its transmission, and the transmission rate was divided into four cases in the simulation process: \( \beta_E = \beta_I = 0.4998; \beta_E = 0.4998/14, \beta_I = 0.4998; \beta_E = 0.4998, \beta_I = 0.4998/14; \beta_E = \beta_I = 0.4998/14. \) The reason for this value here depends on the prevention and control strategy of early isolation for 14 days. The change curve of simulated infected persons is shown in Fig. 9.

Fig. 9 shows that reducing \( \beta_E \) has a greater impact on the spread of the disease than reducing \( \beta_I \). Decreasing \( \beta_I \) has little effect on the peak time of the disease, while decreasing \( \beta_E \) delays the peak time, which is more evident in Fig. 9(b), because it need take some time for the original node to infect the central node. For the betweenness centrality nodes s37, s34, s20 (Fig. 9(a)), decreasing \( \beta_E \) can reduce 9.26% of infected persons, while decreasing \( \beta_I \) can only reduce 4.32% of infected persons. For the original infected nodes in the network (Fig. 9(b)), reducing \( \beta_E \) can reduce the number of infected nodes by 11.45%, while reducing \( \beta_I \) can only reduce the number of infected nodes by 3.85%. This shows the importance of digging out asymptomatic patients to control the spread of the disease, which is closely related to the current policy of timely nucleic acid testing, namely, the timely detection of asymptomatic infected persons and the reduction of transmission.
Fig. 9. The change of the number of symptomatic infected people in the predicted network with time and different transmission rates. a. Betweenness nodes ($s_{37}$, $s_{34}$, $s_{20}$). b. Original nodes ($s_1$, $s_3$, $s_4$).

c. caused by asymptomatic infected persons. In addition, both figures in Fig. 9 show that reducing the two transmission rates at the same time can effectively reduce the number of infected persons by 48.8% and 54.61%, respectively. This means that it makes sense to take all kinds of prevention and control measures to reduce the transmission rate of COVID-19. Also, we can find that at the same time, reducing the transmission rate effect on the spread of the original nodes is greater than the influence of the central nodes. This may be because of the original network node to the path of the central node buffer to infectious center node, so as to make the central node infected probability reduced, and does not cause widespread dissemination around central node.

6. Conclusion and discussion

Up to now, there are still many clusters of confirmed cases of COVID-19 in China, and the situation abroad is grim. Therefore, under the normal situation of prevention and control strategies, how to quickly and effectively control the cluster of epidemic has become our focus. By collecting the data of early confirmed COVID-19 cases in Tianjin, we obtained 136 confirmed COVID-19 cases in Tianjin from January 21, 2020 to March 14, 2020. Baodi District of Tianjin had the most confirmed cases (from the Baodi department store incident), followed by Hedong District and Hebei District. Since the incident in Baodi District department store was an early cluster epidemic, we wanted to provide evidence for the current cluster epidemic control strategy by studying the early Tianjin epidemic.

In order to study the early COVID-19 epidemic in Tianjin, we first constructed a case contact network based on the close contact information of early confirmed cases in Tianjin, in which confirmed cases were nodes and contact between cases were links. Then using the igraph package of R software, the network is visualized and a series of analysis is carried out, and the analysis shows that the network’s cumulative degree distribution roughly obeys the power-law distribution. The results of centrality analysis show that most of the most influential cases in the network are concentrated in Baodi District, followed by Hedong District and Hebei District.

Since the case contact network constructed in this paper may contain false links and missing links, link prediction is needed to reconstruct the network. However, there are many isolated nodes in the network initially constructed. Therefore, assumptions are made to eliminate the isolated nodes in the network, and further link prediction is carried out by using the indexes of CN, AA and RA to reconstruct the network. In terms of the accuracy of the index, RA is the best, followed by AA and CN. According to the prediction result of RA index, the average path of the network is small, and the clustering coefficient is large, which indicates that the predicted case contact network has the nature of a small-world network. The large diameter of the network is a reflection of the wide spread of COVID-19.

Furthermore, we want to know the influence of the central nodes on disease transmission. A susceptible–latent–exposed–infected–recover (SLEIR) model based on individual transmission was established. By simulating disease spread in the original network and the predicted network, we can find the key nodes in the network can accelerate the spread of COVID-19, so identifying the key node is important to control the disease in time. This also reflects that people with larger social circles may accelerate the spread of the disease when the disease spreads through contacts, which corresponds to the spread of COVID-19, and the timely detection and isolation of their multilevel contacts may be reduced if a confirmed case is identified spread of disease. In addition, reducing the transmission rate of the disease can also control the spread of the disease, which requires that the infected people must be isolated, reduce contact with people, which also shows that wearing masks and other prevention and control measures are reasonable. Moreover, reducing the transmission rate of
asymptomatic infected people is more effective in controlling the disease than reducing the transmission rate of infected people, which is consistent with the current policy of detecting positive cases by nucleic acid tests in time.

Because of the occurrence of the cluster epidemic, we still cannot give the strategy to control it quickly. Compared with previous studies, we have given the goal of controlling the spread of COVID-19 from the perspective of the source of infection: detect positive cases in time and reduce the spread of latent. This is consistent with the current goal of multiple rounds of nucleic acid detection. In addition, finding key points quickly and reducing contact transmission rates are powerful ways to control disease transmission. But in real life, at the beginning of the outbreak it is quickly difficult to find the key cases. How to make the key cases in actual life and the key nodes in the network one to one correspondence, we need to address these features in the future.

CRediT authorship contribution statement

Mingtao Li: Conceptualization, Methodology, Modeling, Writing – original draft, Review & editing. Jin Cui: Methodology, Software, Writing – original draft. Juan Zhang: Methodology, Writing – original draft, Review & editing. Xin Pei: Conceptualization, Methodology, Software, Writing – original draft. Guiquan Sun: Supervision, 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.

Data availability

The data that has been used is confidential.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grants (12101443, 12022113, 11801398), Natural Science Foundation of Shanxi Province Grants (20210302124260, 2021030212224095), Preferential Funding for Science and Technology Activities for Returned Overseas Chinese Scholars of Shanxi province (No. 20210009).

References

[1] COVID-19: vaccines, 2022, https://www.who.int/zh/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-(covid-19)-vaccines.
[2] Tracking SARS-CoV-2 variants, 2021, https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.
[3] COVID-19: contact tracing, 2021, https://www.who.int/zh/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-contact-tracing.
[4] X. Ma, X.F. Luo, L. Li, Y. Li, G.Q. Sun, The influence of mask use on the spread of COVID-19 during pandemic in New York City, Results Phys. 34 (2022) 105224, http://dx.doi.org/10.1016/j.rinp.2022.105224.
[5] J.K.K. Asamoah, E. Okyere, A. Abidemi, et al., Optimal control and comprehensive cost-effectiveness analysis for COVID-19, Results Phys. 33 (2022) 105177, http://dx.doi.org/10.1016/j.rinp.2022.105177.
[6] B. Tang, X. Wang, Q. Li, et al., Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, J. Clin. Med. 9 (2) (2020) http://dx.doi.org/10.3390/jcm9020462.
[7] M.T. Li, G.Q. Sun, J. Zhang, et al., Analysis of COVID-19 transmission in Shanxi Province with discrete time imported cases, Math. Biosci. Eng. 17 (4) (2020) 3710–3720, http://dx.doi.org/10.3934/mbe.2020208.
[8] G.Q. Sun, S.F. Wang, M.T. Li, et al., Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources, Nonlinear Dyn. 101 (3) (2021) 1981–1993, http://dx.doi.org/10.1007/s11071-020-05770-9.
[9] M.T. Li, J. Cui, J. Zhang, et al., Transmission analysis of COVID-19 with discrete time imported cases: Tianjin and Chongqing as cases, Infect. Dis. Model. 6 (1) (2021) http://dx.doi.org/10.1016/j.idm.2021.03.007.
[10] G. Xu, Y.Z. Pei, S.Q. Liu, et al., Effectiveness of non-pharmaceutical interventions against local transmission of COVID-19: an individual-based modelling study, Infect. Dis. Model. 6 (2021) 848–858, http://dx.doi.org/10.1016/j.idm.2021.06.005.
[11] J.A. Firth, J. Hellewell, P. Klepac, et al., Using a real-world network to model localized COVID-19 control strategies, Nat. Med. 26 (10) (2020) http://dx.doi.org/10.1038/s41591-020-1036-8.
[12] The official website of Tianjin Health Commission, 2020, http://wsjk.tj.gov.cn/ZTZL1/ZTZL750/YQFKZL9424/YQTB7440/.
[13] Q. Li, X. Guan, P. Wu, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia, N. Engl. J. Med. 382 (13) (2020) http://dx.doi.org/10.1056/NEJMoa2001316.
[14] W.J. Guan, Z.Y. Ni, Y. Hu, et al., Clinical characteristics of coronavirus disease 2019 in China, N. Engl. J. Med. 382 (18) (2020) 1708–1720, http://dx.doi.org/10.1056/NEJMoa2002032.
[15] X.F. Wang, X. Li, G.R. Chen, Network Science: an Introduction, China Higher Education Press, 2012.
[16] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703, http://dx.doi.org/10.1137/070710111.
[17] L.C. Freeman, Centrality in social networks: conceptual clarification, Soc. Netw. 1 (3) (1979) 215–239, http://dx.doi.org/10.1016/0378-8733(79)90021-7#doolink.
[18] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1) (1977) 35–41, http://dx.doi.org/10.2307/3033543.
[19] G. Sabidussi, The centrality index of a graph, Psychometrika 31 (4) (1966) 581–603, http://dx.doi.org/10.1007/BF02289527.
[20] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43, http://dx.doi.org/10.1007/BF02289026.
[21] L.Y. Lu, T. Zhou, Link prediction in complex networks: a survey, Physica A 390 (6) (2011) 1150–1170, http://dx.doi.org/10.1016/j.physa.2010.11.027.

[22] F. Lorrain, H.C. White, Structural equivalence of individuals in social networks, J. Math. Sociol. 1 (1) (1971) 49–80, http://dx.doi.org/10.1080/0022250X.1971.9989788.

[23] L.A. Adamic, E. Adar, Friends and neighbors on the web, Social Networks 25 (3) (2003) 211–230, http://dx.doi.org/10.1016/S0378-8733(03)00009-1.

[24] T. Zhou, L.Y. Lu, Y.C. Zhang, Predicting missing links via local information, Eur. Phys. J. B 71 (4) (2009) 623–630, http://dx.doi.org/10.1140/epjb/e2009-00335-8.

[25] S. Milgram, The small world problem, Psychol. Today 2 (1) (1967) http://dx.doi.org/10.1068/a090185.

[26] D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ networks, Nature 393 (1998) 440–442, http://dx.doi.org/10.1038/30918.

[27] S. Gomez, A. Arenas, J. Borge-Holthoefer, et al., Discrete-time markov chain approach to contact-based disease spreading in complex networks, Europhys. Lett. 89 (3) (2010) http://dx.doi.org/10.1209/0295-5075/89/38009.

[28] X. Pei, Z. Jin, W.Y. Zhang, et al., Detection of infection sources for Avian Influenza A(H7N9) in live poultry transport network during the fifth wave in China, IEEE Access 7 (2019) 155759–155778, http://dx.doi.org/10.1109/ACCESS.2019.2949606.