COVID-19 SIR network modeling and prediction

Yilin Lv [a], Maoting Shi *, Qiao Hu [b]

Northeastern University at Qinhuangdao, Qinhuangdao, China

[a]1608105845@qq.com, * Corresponding Author Email: 2236136493@gmail.com, [b]3202957173@qq.com

Yilin Lv, Maoting Shi and Qiao Hu are co-Author

Abstract. The high infectivity and insidiousness of COVID-19 have enabled the virus to spread rapidly around the world, threatening human life, health and social order, and being infectious in crowded places such as subway stations. Based on the analysis of COVID-19 transmission, this paper divides the population into three categories: susceptible category, discovered disease category and immune category consisting of cure and death, and proposes two models: crowd contact network and contact network-based SIR model to simulate the transmission process of SIR; then absorbs the idea of synchronous update of cellular automaton model, and finally establishes a contact network-based SIR model whose members are affected by all passengers at the same time simulation model. For this model, we fit the data of Beijing, analyze and predict the peak of the latest round of epidemic development (the peak of the number of patients, the peak of the transmission rate), and the length of the latest round of epidemic development cycle, etc, and propose some policy recommendations accordingly.

Keywords: Crowd contact network, SIR network modeling, COVID-19 (Corona Virus Disease 2019).

1. Introduction

At present, new cases are mainly discovered through screening of all employees and active medical treatment. Officials have warned that community transmission has occurred in the local area, and there is a risk of spreading and spilling. The current round of the epidemic [1] revealed that there are still problems in some places, such as weak links such as ideological slackness and paralysis, insufficient emergency preparedness for normalized prevention and control, and inadequate implementation of responsibilities.

Zhong Nanshan, an academician of the Chinese Academy of Engineering, said at the 3rd Chinese Physician Public Welfare Conference that the new crown virus is currently in the active period. Although small-scale clusters of epidemics and sporadic cases have occurred in some areas of my country, they are believed to be available in less than a month. Very effective control. Although the new coronavirus cannot be completely resolved in a short time, it can be effectively controlled.

China calls on other countries around the world to pay attention to the prevention and control of the epidemic and actively carry out local vaccination work. My country is expected to be able to achieve a vaccination rate of 80% and above for urban residents by the end of this year. Vaccination can effectively control the spread of the epidemic.

Zhang Wenhong, director of the Department of Infectious Diseases at Huashan Hospital Affiliated to Fudan University, said that the new crown epidemic cannot be dealt with by doctors alone and requires the cooperation of everyone. Recently, there have been sporadic cases in many places in China. The virus is very hidden. I urge everyone to do nucleic acid testing as soon as possible if they come back from some areas with cases. Rapid virus detection is an important magic weapon to win this epidemic. If it can be detected early, it can be blocked in time, but if it is found late, it may have to pay a greater price. Therefore, it is very important to actively do testing [2].

For the recent new round of the new coronavirus pneumonia epidemic, the above network method is used to simulate the epidemic situation. First, build a contact network [3] between people (or between regions and regions). The data captured from the news of the epidemic report and the official information are used to build a real crowd contact network. The crowd contact network can be represented as a graph G= (V, E), where V is the set of nodes, and E is the set of edges, where each edge is formed by connecting two points in V. If the contact relationship between people is represented as a graph, then the nodes in the graph represent people, and the edges represent the contact relationships between people. Then, combined with the current epidemic prevention measures adopted by our country (such as non-essential travel, nucleic acid detection, and travel code and health code, vaccination), establish SIR by designing the activation mechanism function of network nodes. The network propagation model dynamically tracks the number of susceptible, infected and recovered. Simulate the changes of different groups of people in each cycle (hours, days or weeks). Finally, using the established SIR network transmission model, analyze and predict the peak period of the latest round of epidemic development (peak period of the number of patients, peak period of transmission speed), and the cycle length of the latest round of epidemic development, etc. Combining the actual situation in the next 3-6 months, give practical and feasible suggestions for responding to the development of the epidemic.

2. Model Construction

2.1 Category distance definition

This article intends to distinguish different communities [4] by analogy distance, and the data used is from Beijing epidemic situation data. Suppose that when two nodes i and j belong to different categories, the category distance g(i,j) between nodes i and j is a constant a; when i and j belong to the same category, the category distance value is a constant b(a > B). It can be expressed as the following form:

$$\begin{cases} g(i,j)=a, & \text{i and j belong to different categories} \\ g(i,j)=b, & \text{i and j belong to the same categories} \end{cases} \tag{1}$$

The category distance of the network is now defined as the sum of the category distance values between adjacent nodes of the network. Its formal definition is shown in the following formula. Among them, A represents the network adjacency matrix, and A(i,j)=1 represents that there is a chain edge between nodes i and j.

$$L_A=\frac{1}{2}\sum_{i\neq j,A(I,j)=1} g(I,j) \tag{2}$$

As shown in Fig. 1, nodes 1 to 5 belong to category K, and nodes 6 to 9 belong to category Q. In the figure, the distance values a=5 and b=1 respectively, that is, in the categories K and Q, the category distance values between nodes are both 1, and the nodes belonging to categories K and Q respectively, the category distance value between these nodes is 5, The value on the edge of the chain in the network is the category distance value between two adjacent nodes. According to formula (1), the category distance of the network is 23.

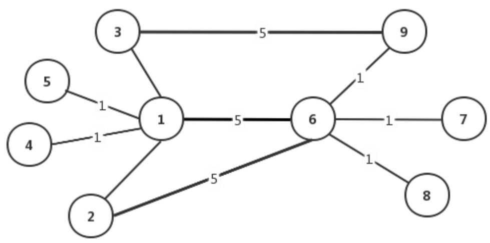

Fig. 1 Network category distance diagram

2.2 Establishment of optimization model

The scale-free attribute [5]is expressed as the optimization goal [6], and the category distance and topological distance of the network are respectively expressed as constraints. The established optimization model is shown in equation (5). Where A is the current evolution network; N is the number of nodes in the network; t is the degree distribution vector of the target network, which is the degree value of node i in the target network; d(A) is the degree distribution vector of network A, di(A) is the degree value of node i in network A; l(A) represents the category distance of network A; z is the category distance of the target network; y is the average shortest topological distance of network A; c is the average shortest topological distance of the target network; F(A) is the scale-free attribute optimization goal of network A; (l(A)-z)2=0 is the category distance constraint;(y-c)2=0 is the topological distance constraint.

$$Min\ F(A) = \sum_{i=1}^{N}(d_i(A) - t_i)^2 \tag{3}$$

$$s.t.\,(l(A) - z)^2 = 0, (y - c)^2 = 0 \tag{4}$$

According to the optimization theory, the Lagrangian relaxation method can be used to transform equation(4) into an unconstrained single-objective optimization problem, that is, equation (4) can be rewritten as shown in equation (5). The θ and δ in formula (5) are Lagrangian multipliers, and the meaning of Lagrangian multipliers is the sensitivity of the objective function to constraint conditions. The values of δ and θ in this article do not affect the final optimization results. There is no need to distinguish the sensitivity of constraint conditions, so directly set θ and δ both to 1. Therefore, equation (5) can be rewritten as equation (6). In this paper, we adopt a greedy algorithm[7] to solve equation (6).

$$\text{Min F'(A)}= \sum_{i=1}^{N} (d_i(A)-t_i)^2+ \theta(l(A)-z)^2+\delta(y-c)^2 \tag{5}$$

$$\text{Min F'(A)}= \sum_{i=1}^{N} (d_i(A)-t_i)^2+ (l(A)-z)^2+(y-c)^2 \tag{6}$$

2.3 Terms, Definitions and formula

Symbolic assumption:

S, I, R-the number of three groups of people at a certain moment

N —Total population

β —Infection rate

α —Elimination rate (cure rate plus death rate)

T —time moment

Differential equation relation:

$$\begin{cases} \frac{dS}{dt}=-\beta\frac{I}{N}S \\ \frac{dI}{dt}=\beta\frac{I}{N}S-\alpha I \\ \frac{dR}{dt}=\alpha I \end{cases} \tag{7}$$

$$N=S+I+R \tag{8}$$

2.4 Assumptions

We assume that people who are cured of the new crown epidemic will not get sick again, that is to say, there are specific antibodies in this body. Because the virus has a too short incubation period, it can even be ignored. When a person is infected with the virus, he becomes an infectious person. In this case, we divide the population into three categories:

The first category: I(t), the number of infected persons at time t. That is, people who are infected with the new crown virus and can transmit the new crown virus to susceptible people.

The second category: S(t), the number of people who are susceptible to infection at time t. People who are not sick but can be infected by the virus.

The third category: R(t), the number of people cured or dead at time t. That is, the number of people who were cured or died after being infected by the virus.

For the spreading rules of the epidemic, we assume:

a. Regardless of the changes in the total population during the period, that is, keep the total population N unchanged.

b. The rate of change of the number of people susceptible to infection S(t) is proportional to the product of I(t) and S(t).

c. The rate of transition from S(t) to R(t) is proportional to S(t).

2.5 State transition in crowd contact networks

In the crowd contact network, the probability of a susceptible person is not only related to β (infection rate), but also to his position in the network[8]. That is, the contact relationship between people is not random, and the probability of each person being in contact with an infected person is no longer equal.

2.5.1 The current node is an infected person

In the next step, the α probability is transformed into an exclusion, and the 1-α probability is still an infected person. As shown in the Fig. 2 below.

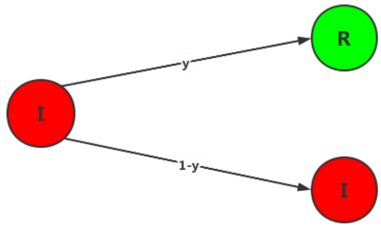

Fig. 2 Crowd contact network (Nodes are infected)

2.5.2 The current node is excluded

In the next step, the probability of 1 is still an excluder. As shown in the Fig. 3 below.

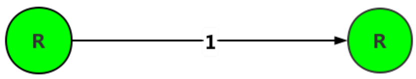

Fig. 3 Crowd contact network (Nodes are excluded)

2.5.3 The current node is susceptible

The status change needs to consider the number of infected persons in neighboring nodes. Assume that there are k infected persons near the node. In the next step, the probability 1-(1-β)2 is transformed into an infected person, and the probability (1-β)2 is still a susceptible person. As shown in the Fig. 4 below.

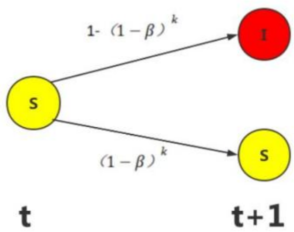

Fig. 4 Crowd contact network (Nodes are susceptible)

3. Model solving and example analysis

3.1 Determination of model parameters

Since the magnitude of the actual uninfected person is quite different from that of the infected person (the susceptible population is of the order of 107 and the infected person does not exceed the order of magnitude), the node transmission method [9] is used, and when the infected population in the node exceeds a certain percentage, the node is defined as being infected, and the actual population data is transformed into node infection status.

Collect more credible data about the new crown epidemic in Beijing, fit the actual data S, I, R, and get the parameters α, β.

Suppose the actual data are S(t0), I(t0) and R(t0), and the fitted data are S(t)、I(t) and R(t), The goal of determining the parameters is to minimize the sum of squares of the total error, that is, the objective function is $\min E=\sum_{t=0}^{n}\left(S(t)\text{-}S(t_0)\right)^2+\left(I(t)\text{-}I(t_0)\right)^2+\left(R(t)\text{-}R(t_0)\right)^2$ .Use the fminsearch function in matlab to solve, and get the parameter α, β when the sum of squares of the total error is the smallest.

The Fig. 5 below shows the change in the minimum variance when the amount of data changes from 3 days to 30 days. It can be seen from the figure that when the amount of data exceeds 22 days, the variance is relatively stable, so the amount of data should be greater than 22 days.

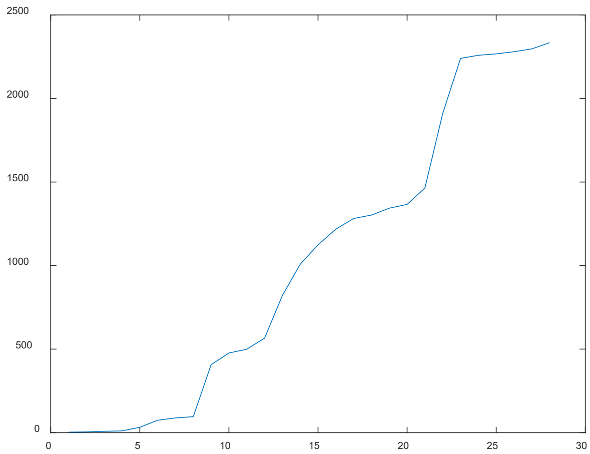

Fig. 5 Minimum variance variation diagram

3.2 Actual fitting

Use the data from the early stage of the outbreak in Beijing, that is, the epidemic data within one month from 2020-1-30 for fitting, parameterα=0.1009, β=0.3012. The obtained image is shown in the Fig. 6 below (where the curve data is the true value, and the point is the fitted value at each time). The image shows that the fit is better. The changes in the numbers of the three groups within a week are shown in the Table 1 below.

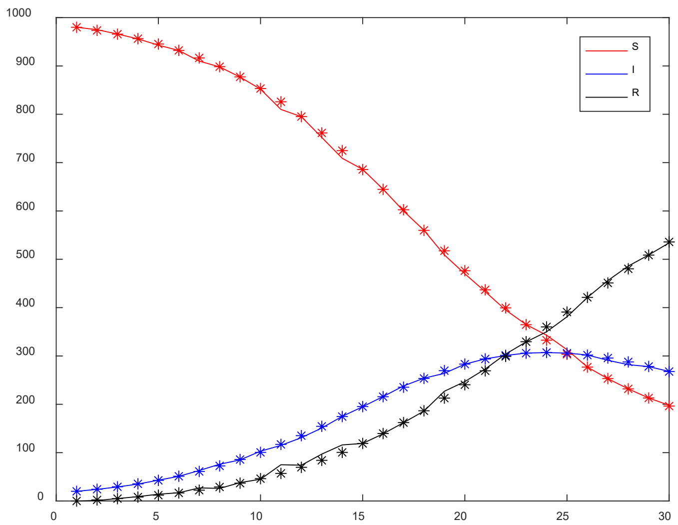

Fig. 6 Fitting diagram

Table 1. Epidemic data

Time	S	I	R
2020-1-30	20	0	980
2020-1-31	24	1	975
2020-2-1	29	5	966
2020-2-2	35	9	956
2020-2-3	43	14	943
2020-2-4	51	17	932
2020-2-5	63	27	910
2020-1-30	20	0	980

Through net logo, you can dynamically observe the evolution of the epidemic. Some of the evolution diagrams are shown in the following Fig. 7 (the red nodes in the figure represent infected persons, blue nodes represent susceptible persons, and gray nodes represent recovered persons).

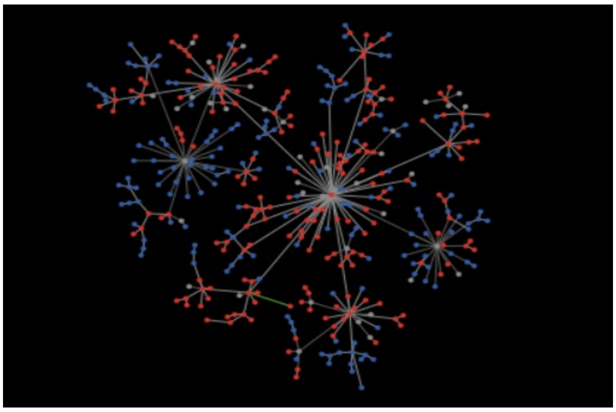

Fig. 7 Epidemic evolution map

Explanatory note: In the figure, the red node represents the infected, the blue node represents the susceptible, and the gray node represents the restorer

The degree of advancement of medical equipment varies between provinces and cities, and the epidemic prevention measures are not the same, which mainly affects the infection rate and recovery rate. The impact of different recovery rates on the epidemic situation is analyzed as shown in the following Fig. 8:

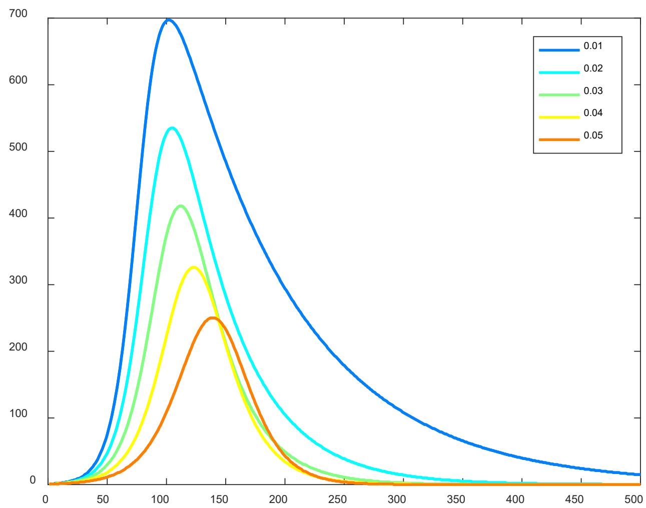

Fig. 8 Effect of different cure rates on epidemic situation

It can be seen from the figure that the recovery rate has an impact on the increase of susceptible persons [10]. The higher the recovery rate, the lower the peak of the increase of susceptible persons, and the earlier the decline and fall back.

4. Conclusions

At present, the foreign epidemic situation is still severe. A new type of super variant Omi Keron has appeared in COVID-19, and the number of foreign infection cases is increasing. The epidemic prevention and control must be normalized, standardized and strict. According to the analysis based on the SIR model in this article, the advanced level of medical equipment and epidemic prevention measures vary among provinces and cities, which mainly affect the infection rate and recovery rate. The recovery rate has an impact on the increase of susceptible people. The higher the recovery rate, the lower the peak increase of susceptible persons, and the earlier the decline and fall back. According to our predictions based on actual data, the number of patients in the new round of epidemic development reached its peak on days 95-141, and the cycle length was about 728 days. Then it is very important to take preventive and control measures.

The first is early detection and early detection. "Discovering the patient" is the first step. If it is difficult to detect the infected person, it will be difficult to carry out the follow-up epidemic prevention and control work. The first step requires a lot of effort to do it. In July 2020, Beijing vigorously contained the rebound of the epidemic in less than a month, which provided us with very good experience. Although the prevention and control efforts are relatively large in the short term, the time window for returning to a normal society is extremely small, and the corresponding cost is the lowest.

Put testing first, maximize testing capabilities in the shortest time, and do everything possible to find and locate patients as quickly and accurately as possible, and follow-up epidemic prevention measures can be followed up immediately. Expanding testing, trying to isolate infected people at home, in shelter, or in hospital, and strictly observe our social distance can effectively cut off the spread of the virus.

The current situation of the world's epidemic situation is extremely uneven, and we should call for an active response to the proposition of how the world can unite and cooperate to fight the epidemic. The new wave of the epidemic has not yet ended, which also sends a very difficult signal to the prevention and control of the epidemic for a long time in the future. It is even more stressed that the risk of epidemic rebound must not be underestimated, and the prevention and control of the epidemic should be divided into regions. For the prevention and control of high-risk areas and high-risk areas, there must be a strict standard, and all key areas must be tested. For high-risk areas and high-risk areas where there is a clear rebound of the epidemic, urban public areas should be temporarily closed, and active prevention and control measures should be taken as soon as possible. This is the experience accumulated in the previous prevention and control process.

In addition, during the fluctuating process of the epidemic, it is necessary to ensure that the local medical resources are sufficient, to avoid overwhelming, and to make targeted compensation enhancements for the conditions of medical facilities in various regions. Areas with advanced levels of prevention and control need assistance. Leading institutions learn from each other, learn from experience, and learn lessons to build a good bridge for epidemic prevention and control.

References

[1] Xu Baochun. SARS infectious disease research based on SIR model [D]. Shandong University, 2019.

[2] Yin Nan. Simulation of the spread of COVID-19 in a limited area based on the SIR model [J]. Statistics and Decision, 2020,36(05):15-20.

[3] Xu Qing, Zu Zhenghu, Zhang Wendou, Xu Zhijing, Huang Peitang, Zheng Tao. Parallel simulation of large-scale population smallpox transmission based on social contact network[J]. Military Medicine,2012,36(10):739-744.

[4] Du Haifeng, Li Shuzhuo, W.F.Marcus, Yue Zhongshan, Yang Xusong.Study on the Community Structure of Small World Networks and Scale-Free Networks[J].Acta Physica Sinica, 2007(12):6886-6893.

[5] Liu Yao. Social Network Feature Analysis and Community Structure Mining[D]. University of Electronic Science and Technology of China, 2013.

[6] Wu Hongrun, Qin Jun, Yi Yunfei, Li Deyi, Zheng Bojin. A community scale-free network model based on optimization theory[J]. Journal of Computers,2015,38(02):337-348.

[7] Xie (Ren Chu), Wang Xiaofan. A review of community structure analysis algorithms in complex networks [J]. Complex Systems and Complexity Science, 2005(03): 1-12.

[8] Luo Bin. Research on Personalized Recommendation Method for Social Networks Based on Power-law Characteristics and Locality-Sensitive Hashing [D]. Beijing Institute of Technology, 2016.

[9] Luo Bin. Research on Personalized Recommendation Method for Social Networks Based on Power-law Characteristics and Locality-Sensitive Hashing [D]. Beijing Institute of Technology, 2016.

[10] Li Chengzhuo, Wu Wentao, Pan Zhenyu, Deng Yujiao, Li Xiao, Dai Zhijun, Lv Jun. Analysis of the prevention and control effect of novel coronavirus pneumonia in Zhejiang Province based on SIR model and basic reproduction number [J]. Zhejiang Medicine, 2020, 42(04):311-314.