Study on the regional risk classification method for the prevention and control of emerging infectious diseases based on directed graph theory

Yong Liu[1], Xiao Wang[1] and Chongqi Zhang[1,2]*

[1]School of Science, Xi'an University of Architecture and Technology, Xi'an, China, [2]School of Economics and Statistics, Guangzhou University, Guangzhou, China

Background: Emerging infectious diseases are a class of diseases that are spreading rapidly and are highly contagious. It seriously affects social stability and poses a significant threat to human health, requiring urgent measures to deal with them. Its outbreak will very easily lead to the large-scale spread of the virus, causing social problems such as work stoppages and traffic control, thereby causing social panic and psychological unrest, affecting human activities and social stability, and even endangering lives. It is essential to prevent and control the spread of infectious diseases effectively.

Purpose: We aim to propose an effective method to classify the risk level of a new epidemic region by using graph theory and risk classification methods to provide a theoretical reference for the comprehensive evaluation and determination of epidemic prevention and control, as well as risk level classification.

Methods: Using the graph theory method, we first define the network structure of social groups and construct the risk transmission network of the new epidemic region. Then, combined with the risk classification method, the classification of high, medium, and low risk levels of the new epidemic region is discussed from two cases with common and looped graph nodes, respectively. Finally, the reasonableness of the classification method is verified by simulation data.

Results: The directed weighted scale-free network can better describe the transmission law of an epidemic. Moreover, the proposed method of classifying the risk level of a region by using the correlation function between two regions and the risk value of the regional nodes can effectively evaluate the risk level of different regions in the new epidemic region. The experiments show that the number of medium and high risk nodes shows no increasing trend. The number of high-risk regions is relatively small compared to medium-risk regions, and the number of low-risk regions is the largest.

Conclusions: It is necessary to distinguish scientifically between the risk level of the epidemic area and the neighboring regions so that the constructed social network model of the epidemic region's spread risk can better describe the spread of the epidemic risk in the social network relations.

KEYWORDS

scale-free network, emerging infectious diseases, graph theory, grading and zoning, epidemic

1. Introduction

Emerging infectious diseases (EIDs) are a class of diseases with a wide range of transmission, multiple modes of transmission, incidence rates much higher than the annual incidence level, difficulty of control, ease of infection of the population, lack of specific treatment and prevention methods (1) and other characteristics. Once an EIDs occurs, it will spread faster and cover a wider area than traditional infectious diseases, causing social problems, affecting the global economy, and even endangering human lives. However, we note that the risk of an epidemic varies from region to region, and people's concerns and social sentiments in different regions also differ. Therefore, a practical classification of the risk level of an epidemic area and the implementation of targeted prevention and control measures according to the risk level are effective means of controlling the rapid spread of the epidemic.

There is a wealth of literature on analyzing the transmission patterns of infectious disease outbreaks and preventative and control measures from various perspectives. On the one hand, studies focus on analyzing and grasping the causes of the occurrence and spread of emerging infectious diseases. For example, Lashley (2) analyzed the factors influencing the occurrence of common emerging infectious diseases (EIDs) and concluded that microbial characteristics are essential in the emergence of infectious diseases. However, humans' behaviors and lifestyle choices are major factors in the emergence and spread of many EIDs. Sabin et al. (3) show how several factors related to human activities play a role in spreading infectious diseases and discuss the main factors contributing to the global spread of the COVID-19 pandemic. Yang and Zhang (4) summarized the uncertainty and complexity of EID and provided the preventive measures for dealing with EIDs.

On the other hand, the model is an effective mathematical tool to analyze and predict infectious disease transmission modes and laws. For example, Chu et al. (5) studied the epidemic spreading in weighted scale-free networks with community structure based on the SI disease model and showed the hierarchical dynamics of the epidemic spreading in the weighted scale-free networks with communities. Sun et al. (6) introduced three modified SIS models on scale-free networks that take into account variable population size, non-linear infectivity, adaptive weights, behavior inertia, and time delay, which could better characterize the actual spread of epidemics. Li et al. (7) using China's prefecture-level high-speed rail network and based on a probabilistic risk model, assessed the risk of COVID-19 infection in 19 provincial-level regions from Wuhan to the whole country in the early stage of domestic transmission, and found that the probability and impact could play different roles in the risk ranking of different regions. Shi et al. (8) developed a comprehensive model for simulating and predicting emerging infectious diseases, based on transmission dynamics and a statistical model driven by public health data. Chowell et al. (9) used the early exponential growth rate method to propose a simple susceptibility-exposure-infection-recovery (SEIR) model, a more complex SEIR model with asymptomatic and hospitalized cases and a stochastic susceptibility-infection-removal (SIR) model with Bayesian estimation to estimate the reproduction number of

Spanish influenza, respectively. Bentout et al. (10) predict the peak time and the number of infectious cases at the peak before and after the implementation of non-pharmaceutical interventions for the COVID-19 based on an age-structured model. Reema et al. (11) compares the effectiveness of the SIR and SEIR models in analyzing epidemic data and also discusses how measures such as social distance and vaccination affect virus transmission in the SIR model.

Mathematical models based on differential equations provide statistical results consistent with the situation in practice, which is an effective tool for studying EIDs. However, as mentioned by Yu and Xue (12), most of the models based on differential dynamical systems are generally computationally cumbersome, and the solutions of the equations are extremely sensitive to the initial conditions, so they cannot deal with the unexpected and random events in the actual process well. In addition, the system formed by the communication, contact, and linkage between people is complex, which leads to the complexity of the transmission process of infectious diseases.

Graph theory is an ideal tool for analyzing, modeling, predicting and forming opinions to formulate strategies to rapidly contain the epidemic and minimize the devastating effects of viral infections (13). Baagyere et al. (14) characterized several complex networks from different domains using concepts from graph theory, and the node degrees, graph spectral radius, degree assortativity, and the entire topological structure of selected complex networks are studied on the SIR epidemic model. It is a fact that graph theory and complex networks are inevitably related. For a complex network, if not considering its dynamic features, the complex network is a graph, and the relevant features of the graph, such as subgraph and complementary graph features, have great significance for the modeling of complex networks (15). In addition, the classical infectious disease model combined with the complex network structure is important for investigating and analyzing the EIDs. Note that a social network (16, 17) is a complex network system based on the relationship between people and established according to certain rules. Zhang et al. (18) constructed an interpersonal network model, and the experimental results indicate that it is feasible and valuable to study virtual social simulation. Moore and Newman (19) studied the transmission characteristics of infectious diseases through the small-world network model and found that the nature of the small world would accelerate the transmission of infectious diseases. Huang et al. (20) found that the scale-free network nodes (21–25) in the complex network satisfy the power law distribution, which is more consistent with the real social network and the transmission law of infectious diseases in the entire society.

In addition, the effective division of regional risk levels is significant for preventing and controlling the rapid spread of EIDs. At present, relevant research has focused on studying regional risk classification methods. For example, Jia et al. (26) constructed an epidemic risk assessment model based on the analysis of population flow data and evaluated the degree of risk for each city using the collected population flow data related to Wuhan, China. Li et al. (27) evaluated the risk level of 38 districts in Chongqing, China, using the single index evaluation method, the analytical hierarchy process, and the systematic clustering method, respectively. Based

on the population migration during COVID-19, Feng et al. (28) constructed a migratory imported risk index by using the number of accumulated cases and the number of new cases and showed that the index could better evaluate the epidemic risk in different places. Tu et al. (29) designed the scoring system with expert consultation and calculated the import, spread, and combined risk scores of regions using quantitative analysis methods to determine the risk level. Using unsupervised machine learning techniques, Fidan et al. (30) applied two clustering methods to classify COVID-19 risk degree.

In this paper, we consider the whole epidemic area as a scale-free network and classify the risk level for each area according to the association between regions. Meanwhile, combining the complex networks with risk assessment models to analyze the risk transmission of infectious diseases process and provide a theoretical basis for evaluating the prevention and control of epidemics and the risk classification of the region. The paper is organized as follows. Section 2 presents the steps for constructing social networks in new epidemic regions. Section 3 presents the method for building the epidemic classification model. Section 4 Simulation of the effectiveness of the method. Section 5 concludes with a brief discussion.

2. Construction of the directed weighted scale-free networks in new epidemic regions

Since infectious disease transmission has been oriented, the directed weighted scale-free networks (31) are more suitable to describe the spread of epidemic risk in real social networks. In the following, we introduce the method of defining the social network as scale-free one and provide the main step for constructing the directed weighted scale-free networks in new epidemic regions.

Let $E \subseteq V \times V$ be the edge set, as introduced by Pastor-Satorras et al. (32), if $V = \{1, 2, \cdots, N\}$ is the node set, then the directed weighted network constructed by N nodes is denoted as $G = (V, E)$.

Now, we introduce the necessary notions as follows. Let k_m be the number of edges connected to the node m. It can be divided into in-degree k_m^{in} and out-degree k_m^{out} for a directed graph, where k_m^{in} is the number of directed edges ending at node m and k_m^{out} is the number of directed edges starting from node m.

In addition, suppose that s_m is the strength of node m, and it can be divided into vertex in weight s_m^{in} and vertex out weight s_m^{out} for a directed graph, where s_m^{in} is the sum of the weights of all edges reaching node m, s_m^{out} is the sum of the weights of all edges starting from node m.

Moreover, let $w_{m,s}$ be the weight of the connection between the node m and s, then the weight of a weighted network is divided into edge weight $w_{m,s}$ and point weight $w_{s,m}$, and the $w_{m,s} \neq w_{s,m}$, $w_{m,s} \neq 0$ if $m \to s$ are connected.

In the following, under the necessary notions above, we will introduce the main steps of the directed weighted scale-free network construction method provided by Barabási et al. (33).

Step 1 *Starting from a network with only two nodes, a new node is added each time and connected to existing ones.*

Step 2 *Assuming that the node m is an existing node, the node with the higher degree is preferentially connected when a new node is added. Suppose that the node s is a new node, the probability of node s connecting to node m is*

$$P(k_s) = \frac{k_s}{\sum\limits_m k_m},$$

where k_s and k_m represent the degree of node s and m, respectively.

Step 3 *Repeat steps 1 and 2 above until the target number of points and edges is reached and the directed weighted scale-free network is constructed.*

3. Method for determining the risk level of epidemic regions

In this section, we will present the methodology for determining the risk level of the epidemic regions.

Now, we assume that the division of regional units is consistent and relatively independent. Let v_m be a node of a new epidemic region. The edge of the network represents the social connection between the two regions, and the weight of the edge represents the correlation between the two regions m and s, usually represented by the correlation function denoted by $L_{m,s}$. Note that an outbreak in one region will affect all connected regions, and the size of $L_{m,s}$ directly determines the size of the epidemic in the region.

3.1. Determination of risk correlation function

To determine the correlation strength $L_{m,s}$ between two regions, we will choose t indicators that affect the $L_{m,s}$ based on the security principle of each region and the relevant personnel and the minimum impact on the economy. In this paper, the five indicators chosen are the distance between two regions, personnel flow, economic traffic, transport convenience, and logistics intensity.

Here we consider the importance of the criteria through the intercriteria correlation (*CRITIC*) method (34) to determine the weight of the indicators t and the correlation function $L_{m,s}$ between the two regions m and s, which the *CRITIC* measures the objective weight of the indicators based on the comparative strength of the evaluation indicators and the conflict between indicators, and the main steps of the method as follows:

Step 1 *Assuming that there are h regions of interest and t indicators that influence the correlation strength, let the matrix consisting of the data of the $i(i = 1, 2, \cdots, t)$th indicator be*

$$\boldsymbol{B}_i = \begin{bmatrix} b_{i(11)} & b_{i(12)} & \cdots & b_{i(1h)} \\ b_{i(21)} & b_{i(22)} & \cdots & b_{i(2h)} \\ \vdots & \vdots & \vdots & \vdots \\ b_{i(h1)} & b_{i(h2)} & \cdots & b_{i(hh)} \end{bmatrix},$$

where $b_{i(ms)}$ is the original data of the indicator i between regions m and s.

Step 2 *Denote*

$$S_i = \begin{bmatrix} s_{i(11)} & s_{i(12)} & \cdots & s_{i(1h)} \\ s_{i(21)} & s_{i(22)} & \cdots & s_{i(2h)} \\ \vdots & \vdots & \vdots & \vdots \\ s_{i(h1)} & s_{i(h2)} & \cdots & s_{i(hh)} \end{bmatrix},$$

where $s_{i(ms)}$ is the index value of $b_{i(ms)}$ by dimensionless (35). If the size of $b_{i(ms)}$ is proportional to the risk, we have

$$s_{i(ms)} = \frac{b_{i(ms)} - \min(b_{i(ms)})}{\max(b_{i(ms)}) - \min(b_{i(ms)})},$$

otherwise,

$$s_{i(ms)} = \frac{\max(b_{i(ms)}) - b_{i(ms)}}{\max(b_{i(ms)}) - \min(b_{i(ms)})}.$$

Step 3 *Let*

$$\overline{s_{i(ms)}} = \frac{1}{hh \sum_{s=1}^{hh} s_{i(ms)}}.$$

Then, the standard deviation of the i-th index is given by

$$P_i = \sqrt{\frac{\sum_{s=1}^{hh} \left(s_{i(ms)} - \overline{s_{i(ms)}}\right)^2}{hh - 1}}.$$

Step 4 *Let R_i be a measure of the conflict created by the j-th indicator with respect to the i-th indicator, we have*

$$R_i = \sum_{j=1}^{t} \left(1 - r_{ij}\right), \qquad (1)$$

where r_{ij} represents the correlation coefficient between the evaluation index i and j.

Step 5 *Denote*

$$C_i = P_i \sum_{j=1}^{t} \left(1 - r_{ij}\right) = P_i R_i.$$

Then, the objective weight is defined as

$$w_i = \frac{C_i}{\sum_{i=1}^{t} C_i}.$$

Step 6 *The weighted standardized matrix $R = \left(r_{i(ms)}\right)_{t(hh)}$ is obtained, where $r_{i(ms)} = w_i s_{i(ms)}$, w_i is the index weight determined by the above steps.*

Following the above steps, the correlation function $L_{m,s}$ between the two regions m and s is

$$L_{m,s} = \sum_{i=1}^{t} w_i * s_{i(ms)} \qquad (2)$$

3.2. Calculation of node risk function and risk value

Let p be the value of the risk of the epidemic occurring in a region of the social network. If the epidemic occurs in this region, $p = 1$, while for other regions (nodes), $p = 0$. We first define the nodes that are connected to the nodes in layer 1 (in addition to the defined nodes) as layer 2 nodes, and so on until we have defined all the nodes in the network. Note that if an infectious disease occurs in a node (region), another node directly connected to that node will be the first affected node, i.e., a layer 1 node. In particular, if the node is the first node in the scale-free network, the risk value is 1.

Furthermore, we define the direction of the edges as the direction of the epidemic spread and construct a weighted scale-free network for the spread of risk across the epidemic region. Since scale-free networks have cyclic graphs (36) and are calculated differently from ordinary nodes. In the following, we focus on two different node cases to provide methods for computing the risk function and risk value, respectively.

Case 1: Calculation of risk value of common node

Now, starting from the node whose risk value is 1 (the initial node), we calculate the risk value of the first layer node connected to it. According to the direction of risk transmission, if the in-degree of node s is 1, it shows that node s is influenced by a node. Assuming that node s is influenced by node m, the risk value of node s can be defined as the product of the risk value of node m and the association function between two nodes as follows:

$$p_s = p_m L_{m,s}, \qquad (3)$$

where p_s is the risk value of the node s and p_m is the risk value of the node m.

Let the set $\pi(s_n)$ represents the set of all other nodes that affect the node s, then the risk value can be defined as the sum of the risk value and the correlation function product of the node s and all the nodes s_n on $\pi(s_n)$, that is

$$p_s = \sum_{s_n \in \pi_{(s_n)}} p_{s_n} l_{s_n,s} L_{s_n,s}, \qquad (4)$$

where

$$l_{s_n,s} = \frac{L_{s_n,s}}{\sum_{s_n \in \pi_{(s_n)}} L_{s_n,s}}.$$

Therefore, starting from each node where we have obtained the risk value, we find the nodes connected to it and then calculate the risk value of all nodes in the obtained scale-free network according to Equations (3) and (4).

Case 2: Calculation of risk value of the loop graph node

Note that the risk transmission direction of nodes at different levels is transferred from the upper node to the lower node, the risk

TABLE 1 Risk classification criteria.

Risk level	High-risk	Medium-risk	Low-risk
Value at risk	$p \geq p_{s_1}$	$p_{s_2} \leq p < p_{s_1}$	$p < p_{s_2}$

transmission is unidirectional. If all the nodes are only influenced by their upper-layer nodes, the risk value of the node is only related to the upper-layer nodes, and the node does not influence the risk value of the upper-layer nodes.Therefore, the method of calculating the risk value of each node is the same as that of the common nodes, and the Equations (3) and (4) can be used directly to calculate the risk values of nodes in different layers. However, it should be noted that for a loop graph node there is no order between them that can influence each other, and the risk transmission direction is bidirectional when the nodes of the same layer are connected, and Equations (3) and (4) cannot be used.

3.2.1. Classification of risk levels

Using the same method of the three-level risk rating system, as introduced in Tu et al. (37), we have the classification criteria as shown in Table 1.

It follows that the risk classification criteria given in Table 1, we need to calculate the risk value of each node to accurately classify the risk level of the area represented by all nodes. Then, according to Table 1, the whole area can be classified into high risk, medium risk and low risk, and the corresponding precise prevention and control measures can be taken for the different risk level of each region.

TABLE 2 The correlation function for group 1.

$L_{1,2} = 0.7117$	$L_{1,3} = 0.3215$	$L_{1,4} = 0.4431$	$L_{1,5} = 0.6580$		
$L_{2,22} = 0.8179$	$L_{2,23} = 0.3893$	$L_{2,24} = 0.2132$	$L_{2,25} = 0.8065$		
$L_{3,16} = 0.4333$	$L_{3,17} = 0.8896$	$L_{3,18} = 0.7086$	$L_{3,19} = 0.0171$	$L_{3,20} = 0.1932$	$L_{3,21} = 0.8897$
$L_{4,6} = 0.9380$	$L_{4,7} = 0.6014$				
$L_{5,11} = 0.1752$	$L_{5,12} = 0.1096$	$L_{5,13} = 0.7432$	$L_{5,14} = 0.2404$	$L_{5,15} = 0.6849$	
$L_{8,22} = 0.8328$	$L_{8,47} = 0.1397$	$L_{8,48} = 0.7896$	$L_{8,49} = 0.1932$	$L_{8,50} = 0.4378$	
$L_{9,22} = 0.7266$					
$L_{10,22} = 0.3939$					
$L_{13,14} = 0.9448$					
$L_{23,35} = 0.7992$	$L_{23,36} = 0.6164$	$L_{23,37} = 0.2802$	$L_{23,38} = 0.0652$	$L_{23,39} = 0.8393$	$L_{23,40} = 0.5297$
$L_{24,31} = 0.9806$	$L_{24,32} = 0.0496$	$L_{24,33} = 0.9852$	$L_{24,34} = 0.3359$		
$L_{25,26} = 0.9701$	$L_{25,27} = 0.8291$	$L_{25,28} = 0.6448$	$L_{25,29} = 0.2832$	$L_{25,30} = 0.5597$	
$L_{38,41} = 0.6088$	$L_{38,42} = 0.0043$	$L_{38,43} = 0.2152$	$L_{38,44} = 0.5119$	$L_{38,45} = 0.8964$	$L_{38,46} = 0.6535$
$L_{42,43} = 0.2537$					
$L_{48,49} = 0.8281$					

TABLE 3 The correlation function for group 2.

$L_{1,2} = 0.8147$	$L_{1,3} = 0.2760$	$L_{1,4} = 0.1622$	$L_{1,5} = 0.4173$		
$L_{2,22} = 0.6433$	$L_{2,23} = 0.9361$	$L_{2,24} = 0.0596$	$L_{2,25} = 0.3015$		
$L_{3,16} = 0.4299$	$L_{3,17} = 0.9160$	$L_{3,18} = 0.5822$	$L_{3,19} = 0.7363$	$L_{3,20} = 0.8507$	$L_{3,21} = 0.6797$
$L_{4,6} = 0.7943$	$L_{4,7} = 0.0479$				
$L_{5,11} = 0.3789$	$L_{5,12} = 0.5468$	$L_{5,13} = 0.6802$	$L_{5,14} = 0.7011$	$L_{5,15} = 0.0942$	
$L_{8,22} = 0.5407$	$L_{8,47} = 0.3947$	$L_{8,48} = 0.5606$	$L_{8,49} = 0.9448$	$L_{8,50} = 0.3112$	
$L_{9,22} = 0.9027$					
$L_{10,22} = 0.0714$					
$L_{13,14} = 0.9448$					
$L_{23,35} = 0.6433$	$L_{23,36} = 0.0376$	$L_{23,37} = 0.8116$	$L_{23,38} = 0.6663$	$L_{23,39} = 0.5985$	$L_{23,40} = 0.4624$
$L_{24,31} = 0.8699$	$L_{24,32} = 0.6834$	$L_{24,33} = 0.9296$	$L_{24,34} = 0.9457$		
$L_{25,26} = 0.5328$	$L_{25,27} = 0.2316$	$L_{25,28} = 0.0714$	$L_{25,29} = 0.52391$	$L_{25,30} = 0.4709$	
$L_{38,41} = 0.4243$	$L_{38,42} = 0.2684$	$L_{38,43} = 0.7040$	$L_{38,44} = 0.6967$	$L_{38,45} = 0.3507$	$L_{38,46} = 0.4889$
$L_{42,43} = 0.0516$					
$L_{48,49} = 0.6981$					

4. Experiments and simulation

In this section, we aim to verify the feasibility of the regional risk classification method proposed in this paper by simulation. Firstly, we use *MATLAB* software to construct a scale-free network with nodes that follow a power law distribution and have the characteristics of a real network. Secondly, we

consider five necessary indicators that influence the correlation intensity of regional risk, namely the distance between two regions, personnel flows, economic traffic, transport convenience and logistics intensity, and use the *MATLAB* software to generate two sets of original data for each of the above five indicators, as shown in the Supplementary material. Moreover, using Equation (2), we calculate the correlation intensity of two sets of original data for
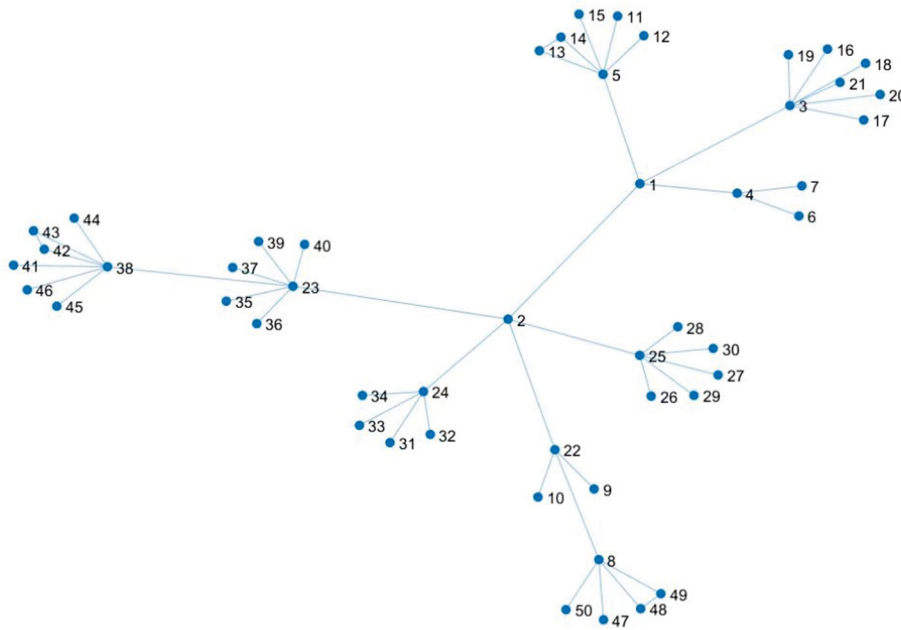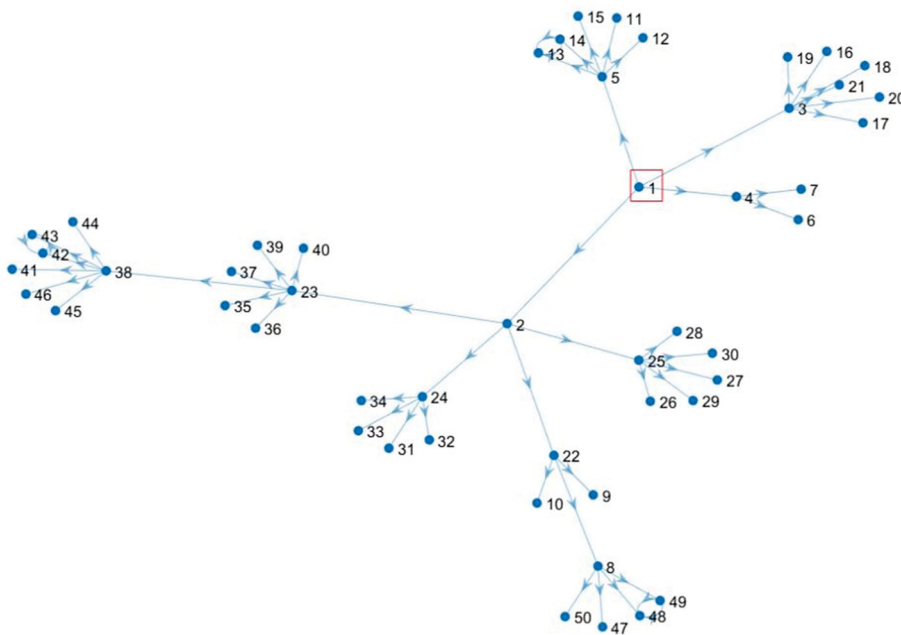

FIGURE 1
Scale-free social networks.

FIGURE 2
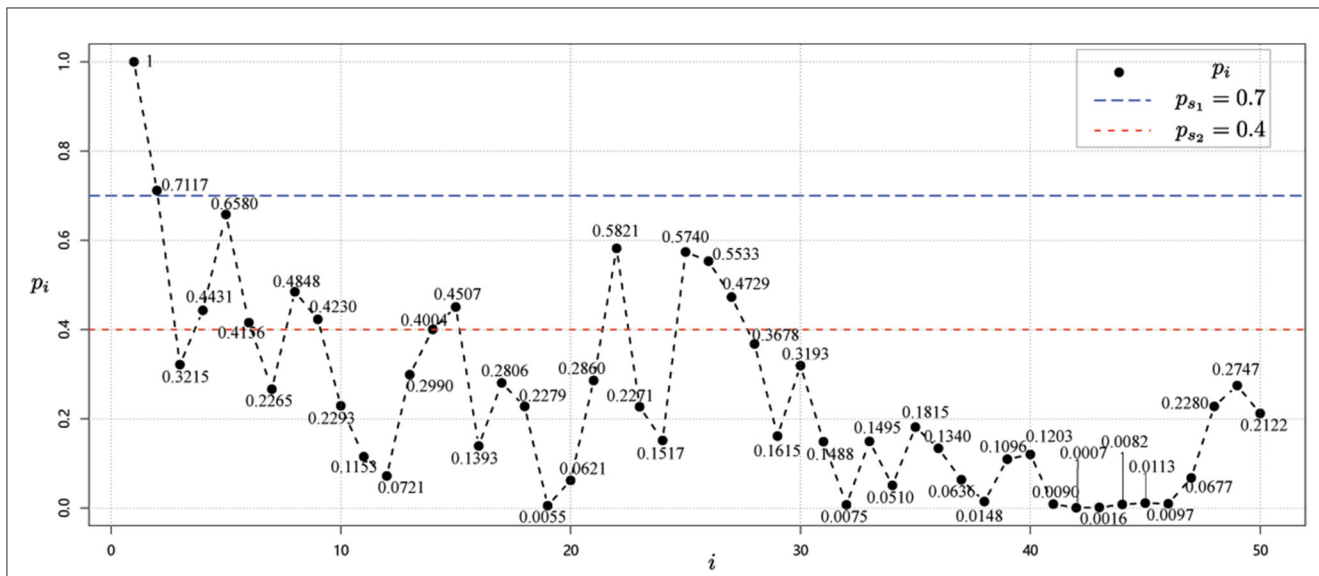Directed scale-free social network during v_1 outbreak.

FIGURE 3
Risk value of each node $i, i = 1, 2, \cdots, 50$ for data 1 for case 1.

TABLE 4 The risk grade classification of each node for data 1 for case 1.

Risk levels	High-risk	Medium-risk	Low-risk
Region	v_1, v_2	$v_4, v_5, v_6,$	$v_3, v_7, v_{10}, v_{11}, v_{12},$ $v_{13},$
		$v_8, v_9, v_{14},$	$v_{16}, v_{17}, v_{18}, v_{19}, v_{20},$ $v_{21},$
		$v_{15}, v_{22}, v_{25},$	$v_{23}, v_{24}, v_{28}, v_{29}, v_{30},$ $v_{31},$
		v_{26}, v_{27}	$v_{32}, v_{33}, v_{34}, v_{35}, v_{36},$ $v_{37},$
			$v_{38}, v_{39}, v_{40}, v_{41}, v_{42},$ $v_{43},$
			$v_{44}, v_{45}, v_{46}, v_{47}, v_{48},$ v_{49}, v_{50}

these five indicators as shown in Tables 2, 3 below. Finally, we analyzed the two sets of experimental data using the presented classification method of regional risk levels.

4.1. Construction of a scale-free social network

Using the *MATLAB* software, we construct the scale-free network as shown in Figure 1. There are 50 nodes, whose node distribution corresponds to the characteristics of the scale-free network and the actual new epidemic regions.

4.2. Experimental analysis

Due to the scale-free social network constructed in the simulation, there is more than one node in the new outbreak

area. Therefore, we will discuss the proposed method in the whole network with epidemics at one and two nodes respectively. At the same time, two sets of correlation strength data are used for experimental comparison and analysis, which makes the simulation experiment more effective.

- **Experiment 1**

Experiment 1 focuses on Group 1 data and discusses the construction of a regional outbreak risk transmission network when an outbreak occurs at one node and two nodes in a scale-free network, and the risk level of the whole region is classified using the proposed risk level classification method.

Case 1: An outbreak has occurred in a region.

Assuming the epidemic occurs in v_1, the risk value of the v_1 node is 1, and the resulting directed scale-free network is shown in Figure 2.

Next, starting from the node v_1 whose risk value is 1, and using the correlation strength in Table 2, we can calculate the risk values for each node $i, i = 1, 2, \cdots, 50$ as shown in Figure 3.

Using the method that provided in Equations (3) and (4), the risk classification values are $p_{s_1} = 0.7$ and $p_{s_2} = 0.4$. Then, we can then obtain the risk level classification of each node from the risk values and the risk class classification criterion (37). As shown in Figure 3, it is easy to see that the risk level classification of each node is as given in Table 4.

Case 2: Outbreaks occurred in two regions.

In the case 2, we focus on the data of Group 1 and discuss the construction of a regional epidemic risk transmission network when an epidemic occurs at two nodes in a scale-free network, and the risk level of the whole region is classified using the proposed risk level classification method.

Suppose the epidemic occurs at node v_1 and node v_{22}, then the risk value of node v_1 and node v_{22} is 1, and the formed directed scale-free network is shown in Figure 4. Starting from node v_1 and node v_{22} whose risk value is 1, we calculate the risk value of other nodes by considering the correlation strength in data 1, which is shown in Figure 5 below.

Then, according to the calculated risk values, we have the risk grade classification of each node, which is shown in Table 5 below.

● **Experiment 2**

In Experiment 2, we consider the same method as in Experiment 1. For the data in Group 2, the classification of the regional risk level in the constructed scale-free social network is considered for the cases where the outbreak occurs in one node and in two nodes, respectively.
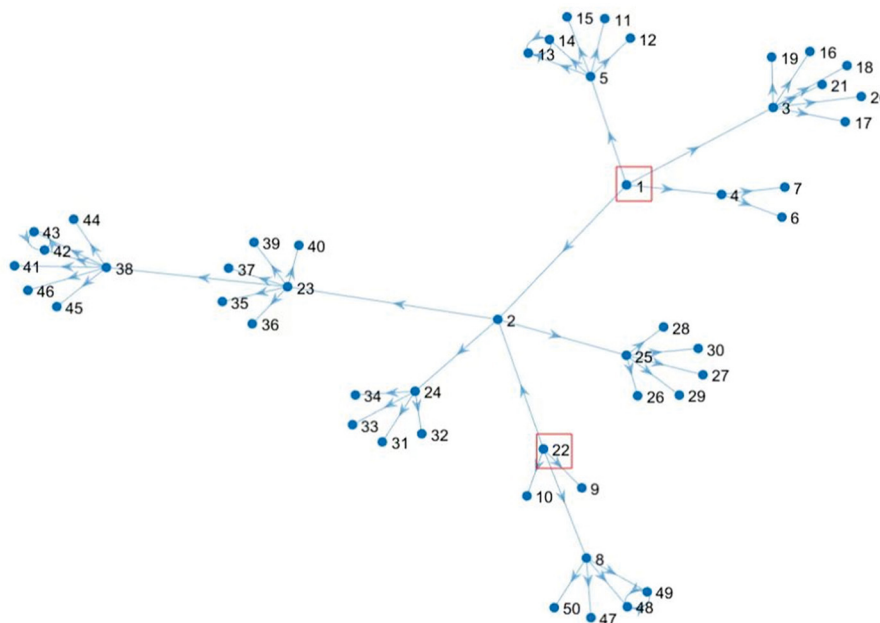
Case 1: An outbreak has occurred in one region.

FIGURE 4
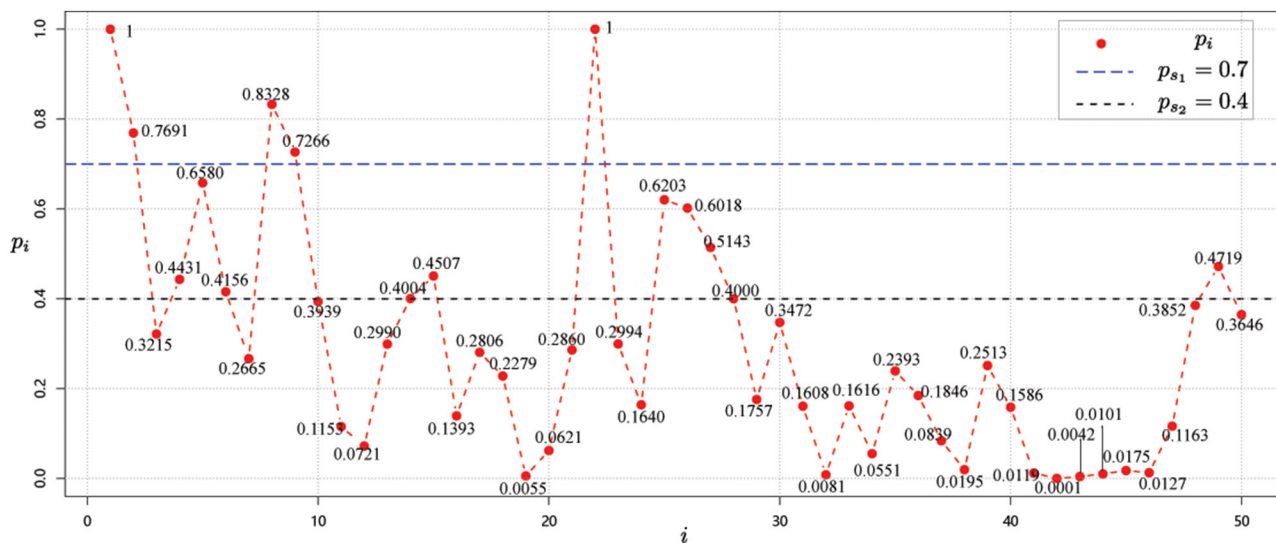Directed scale-free social network during v_1 and v_{22} outbreaks.

FIGURE 5
Risk value of each node i, $i = 1, 2, \cdots, 50$ for data 1 for case 2.

Now, considering the scale-free network as shown in Figure 2 for case 1. Starting from the node v_1 with risk value 1, we get the risk values of other nodes by combining the association strength calculation in Group 2. Then, we get the risk value of each node $i, i = 1, 2, \cdots, 50$ as shown in Figure 6.

Thus, based on the risk values obtained from the red dots in Figure 6 and the risk level classification guidelines in Table 1, we obtain the following risk level classification for each node.

- **High-risk regions:** v_1, v_2, v_{22}, v_{23};
- **Medium-risk regions:** $v_5, v_9, v_{37}, v_{38}, v_{39}, v_{42}, v_{49}$;
- **Low-risk regions:** $v_3, v_4, v_6, v_7, v_8, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{24}, v_{25}, v_{26}, v_{27}, v_{28}, v_{29}, v_{30}, v_{31}, v_{32}, v_{33}, v_{34}, v_{35}, v_{36}, v_{40}, v_{41}, v_{43}, v_{44}, v_{45}, v_{46}, v_{47}, v_{48}, v_{50}$.

Case 2: Outbreaks occurred in two regions

The situation is the same as shown in Figure 4. Next, starting with node v_1 and node v_{22}, whose risk value is 1, the risk value of other nodes is calculated for the correlation strength in Group 2, as shown by the green dots in Figure 6.

TABLE 5 The risk grade classification of each node for data 1 for case 2.

Risk levels	High-risk	Medium-risk	Low-risk
Region	$v_1, v_2,$	$v_4, v_5, v_6, v_{14},$	$v_3, v_7, v_{10}, v_{11}, v_{12}, v_{13},$
	$v_8, v_9,$	$v_{15}, v_{25}, v_{26},$	$v_{16}, v_{17}, v_{18}, v_{19}, v_{20},$
	v_{22}	v_{27}, v_{28}, v_{49}	$v_{21}, v_{23}, v_{24}, v_{29}, v_{30},$
			$v_{31}, v_{32}, v_{33}, v_{34}, v_{35},$
			$v_{36}, v_{37}, v_{38}, v_{39}, v_{40},$
			$v_{41}, v_{42}, v_{43}, v_{44}, v_{45},$
			$v_{46}, v_{47}, v_{48}, v_{50}$

By considering the risk level classification guidelines in Table 1, we have

- **High-risk regions:** $v_1, v_2, v_9, v_{22}, v_{23}$;
- **Medium-risk regions:** $v_5, v_8, v_{35}, v_{37}, v_{38}, v_{39}, v_{42}, v_{49}$;
- **Low-risk regions:** $v_3, v_4, v_6, v_7, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{24}, v_{25}, v_{26}, v_{27}, v_{28}, v_{29}, v_{30}, v_{31}, v_{32}, v_{33}, v_{34}, v_{36}, v_{40}, v_{41}, v_{43}, v_{44}, v_{45}, v_{46}, v_{47}, v_{48}, v_{50}$.

5. Conclusion

This paper uses graph theory and the risk assessment method to construct the epidemic risk classification method. In addition, the rationality and effectiveness of the classification method are verified by simulation. Furthermore, we use the MATLAB software to construct a scale-free network and generate the original the original data of the five required indicators. Then the risk classification degree in each node for two cases of the epidemic occurring in one node and two nodes are discussed. The experiment shows that the number of medium and high risk nodes does not show a significant increasing trend, and the number of high risk regions is relatively small compared to the number of medium risk regions, and the number of low risk regions is the largest, which is consistent with the classification of regional risk levels in the real society.

The construction of the social network of risk classification in new epidemic regions by the directed weighted scale-free network is more suitable for the transmission law of epidemic occurrence risk. It describes the transmission status of epidemic occurrence risk in social network relations. Furthermore, the established regional risk classification method can well classify the risk levels of different regions in the new epidemic area by determining the correlation function between the two regions and the risk value of the regional node. The experiment verified the rationality of
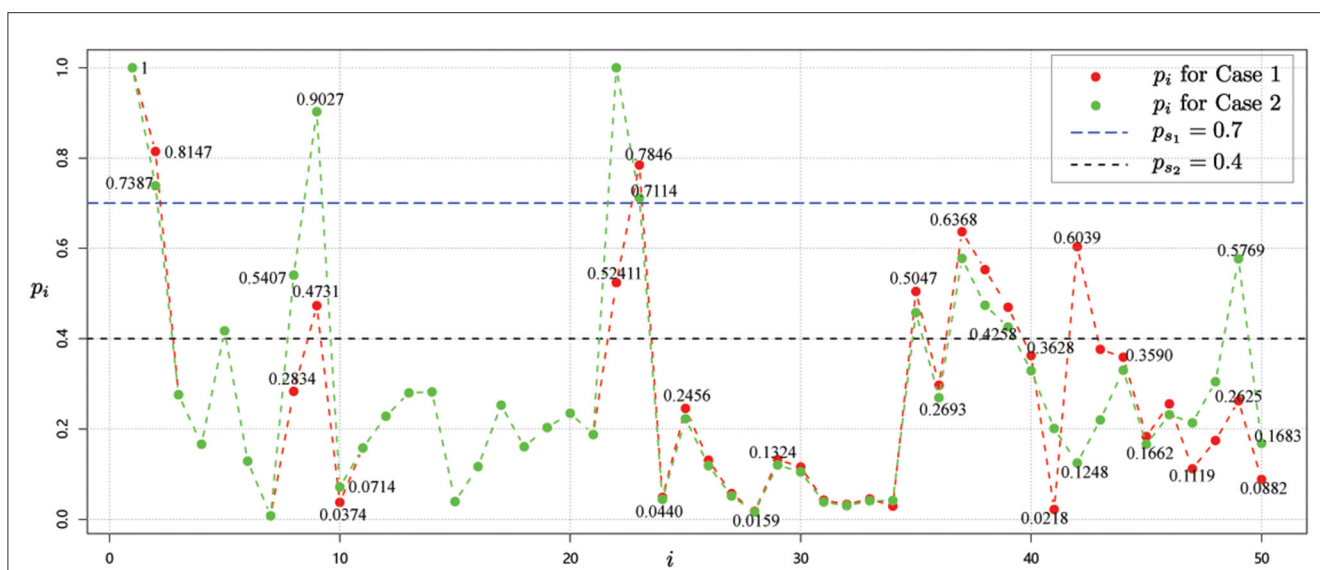

FIGURE 6
Risk value of each node $i, i = 1, 2, \cdots, 50$ for data 2 for case 1 and case 2.

the method, and it can provide a theoretical basis for the government to quickly judge the risk levels of different regions in epidemic prevention.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

XW: formal analysis and writing—review and editing. YL: methodology. CZ: resources. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1211291/full#supplementary-material

References

1. Ding LR, Li Y. Research on spatial and temporal differentiation of public opinion in the situation of major infectious diseases. *J Mod Inform.* (2023) 43:120–31. doi: 10.3969/j.issn.1008-0821.2023.01.012

2. Lashley FR. Factors contributing to the occurrence of emerging infectious diseases. *Biol Res Nurs.* (2003) 4:258–67. doi: 10.1177/1099800403251238

3. Sabin NS, Calliope AS, Simpson SV, Arima H, Ito H, Nishimura T, et al. Implications of human activities for (re)emerging infectious diseases, including COVID-19. *J Physiol Anthropol.* (2020) 39:1–12. doi: 10.1186/s40101-020-00239-5

4. Yang W, Zhang, T. Strategy and measures in response to highly uncertain emerging infectious disease. *Chin J Epidemiol.* (2022) 43:627–33. doi: 10.3760/cma.j.cn112338-20220210-00106

5. Chu X, Guan J, Zhang Z, Zhou S. Epidemic spreading in weighted scale-free networks with community structure. *J Stat Mech.* (2009) 2009:P07043. doi: 10.1088/1742-5468/2009/07/P07043

6. Sun M, Zhang H, Kang H, Zhu G, Fu X. Epidemic spreading on adaptively weighted scale-free networks. *J Math Biol.* (2017) 74:1263–98. doi: 10.1007/s00285-016-1057-6

7. Li T, Rong L, Zhang A. Assessing regional risk of COVID-19 infection from Wuhan via high-speed rail. *Transport Policy.* (2021) 106:226–38. doi: 10.1016/j.tranpol.2021.04.009

8. Shi XL, Wei FF, Chen WN. A swarm-optimizer-assisted simulation and prediction model for emerging infectious diseases based on SEIR. *Complex Intell Syst.* (2022) 9:2189–204. doi: 10.1007/s40747-022-00908-1

9. Chowell G, Nishiura H, Bettencourt LM. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J R Soc Inter.* (2007) 4:155–66. doi: 10.1098/rsif.2006.0161

10. Bentout S, Tridane A, Djilali S, Touaoula TM. Age-Structured modeling of COVID-19 epidemic in the USA, UAE and Algeria. *Alex Eng J.* (2021) 60:401–11. doi: 10.1016/j.aej.2020.08.053

11. Reema G, Babu B, Tumuluru P, Praveen S. COVID-19 EDA analysis and prediction using SIR and SEIR models. *Int J Healthcare Manage.* (2022) 1–16. doi: 10.1080/20479700.2022.2130630

12. Yu L, Xue HF. Epidemic spread model based on complex networks. *J Shanxi Univ Sci Technol.* (2007) 25:126–9.

13. Rad D, Venkatraman Y, Krithicaa NY, Balas VE. Graph theory applications to comprehend epidemics spread of a disease. *Brain Broad Res Artif.* (2021) 12:161–77. doi: 10.18662/brain/12.2/198

14. Baagyere EY, Qin Z, Xiong H, Qin Z. Characterization of complex networks for epidemics modeling. *Wireless Pers Commun.* (2015) 83:2835–58. doi: 10.1007/s11277-015-2569-x

15. Duan ZS. Graph theory and complex networks. *Adv Mech.* (2008) 38:702–12. doi: 10.6052/1000-0992-2008-6-J2008-083

16. Dekker A. Applying social network analysis concepts to military C4ISR architectures. *Connections.* (2002) 24:93–103.

17. Wasserman S. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press (1994). doi: 10.1017/CBO9780511815478

18. Zhang MZ, Yu YY, Hu XF, Si GY. Study on social network's modeling virtual social simulation. *Comput Simul.* (2009) 26:14–7. doi: 10.3969/j.issn.1006-9348.2009.02.005

19. Moore C, Newman ME. Epidemics and percolation in small-world networks. *Phys Rev E.* (2000) 61(5 Pt B): 5678–82. doi: 10.1103/PhysRevE.61.5678

20. Huang Q, Min LQ, Chen X. Susceptible-infected-recovered models with natural birth and death on complex networks. *Math Method Appl Sci.* (2015) 38:37–50. doi: 10.1002/mma.3048

21. Barabasi A, Albert R. Emergence of scaling in random networks. *Science.* (1999) 286:509–12. doi: 10.1126/science.286.5439.509

22. Goh KI, Kahng B, Kim D. Universal behavior of load distribution in scale-free networks. *Phys Rev Lett.* (2001) 87:278701. doi: 10.1103/PhysRevLett.87.278701

23. Newman ME. The structure and function of complex networks. *SIAM Rev.* (2003) 45:167–256. doi: 10.1137/S003614450342480

24. Del-Genio CI, Gross T, Bassler, KE. All scale-free networks are sparse. *Phys Rev Lett.* (2011) 107:178701. doi: 10.1103/PhysRevLett.107.178701

25. Courtney OT, Bianconi G. Dense power-law networks and simplicial complexes. *Phys Rev E.* (2018) 97:052303. doi: 10.1103/PhysRevE.97.052303

26. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature.* (2020) 582:389–94. doi: 10.1038/s41586-020-2284-y

27. Li HZ, Yi DL, Li GM. Comparative study on assessment methods of epidemic risk of COVID-19 in Chongqing. *J Chongqing Med Univ.* (2020) 45:870–75. doi: 10.13406/j.cnki.cyxb.002564

28. Feng ZX, Duan DY, Liu Y. Risk assessment of the COVID-19 and comparison of response strategies. *J Xi'an Univ Technol.* (2020) 36:439–46. doi: 10.19322/j.cnki.issn.1006-4710.2020.04.001

29. Tu YW, Zhong RX, Xiao JP, Xiao LY, Li Y, Song L. Region- and risk-specific strategies for coronavirus disease 2019 epidemic control and prevention in Guangdong province: a risk assessment study. *Chin J Publ Health.* (2020) 36:486–92. doi: 10.11847/zgggws1128921

30. Fidan HG, Yüksel ME.A comparative study for determining COVID-19 risk levels by unsupervised machine learning methods. *Expert Syst Appl.* (2021) 190:116243. doi: 10.1016/j.eswa.2021.116243

31. Pi XC, Tang LK, Chen XZ. A directed weighted scale-free network model with an adaptive evolution mechanism. *Phys A.* (2021) 572:125897. doi: 10.1016/j.physa.2021.125897

32. Pastor-Satorras R, Castellano C, Van Mieghem, P, Vespignani A. Epidemic processes in complex networks. *Rev Mod Phys.* (2015) 87:925–79. doi: 10.1103/RevModPhys.87.925

33. Barabási AL, Ravasz E, Vicsek T. Deterministic scale-free networks. *Phys A.* (2001) 299:559–64. doi: 10.1016/S0378-4371(01)00369-7

34. Diakoulaki D, Mavrotas G, Lefteris P. Determining objective weights in multiple criteria problems: the CRITIC method. *Comput Oper Res.* (1995) 22:763–70. doi: 10.1016/0305-0548(94)00059-H

35. Li, WW, Yi PT, Liu, LY. Outliers recognition and the dimensionless method in comprehensive evaluation. *Oper Res Manage Sci.* (2018) 27:173–8. doi: 10.12005/orms.2018.0099

36. Sinha D, Kaur B, Zaslavsky T. The characteristic polynomial of a graph containing loops. *Discrete Appl Math.* (2021) 300:97–106. doi: 10.1016/j.dam.2021.04.025

37. Tu HW, Zhong RX, Xiao JP, et al. Research on the prevention and control strategy of novel coronavirus pneumonia in Guangdong Province by zoning and grading. *China Public Health.* (2020) 36:486–92.