https://doi.org/10.26599/IJCS.2022.9100027 International Journal of Crowd Science

Individual Behavior Modeling and Transmission Control During Disease Spread: A Review

Wengxiang Dong[1] ✉ and H. Vicky Zhao[1]

ABSTRACT

In this paper, we provide a detailed review of two categories of the literature: the spontaneous protective behaviors of individuals during disease spread and the mandatory measures to control the disease spread. In the literature, the models of individual protective behaviors can be divided into two parts: the environment-induced protective behaviors and the information-induced protective behaviors. And the mandatory measures of disease control can be divided into two parts: the macro-based control methods and the micro-based control methods. We provide a detailed review to the various categories of research. Then we compare the effects of different control methods through simulation. Among the micro-based control methods, the method based on minimizing the largest eigenvalue has the best effect. This review is of crucial importance to summarize the studies of the spontaneous protective behaviors during disease spread and the mandatory measures to control the disease spread.

KEYWORDS

individual behavior modeling; disease spread control; crowd networks

The recent outbreak of COVID-19 has resulted in global-scale infections and huge economic losses. As of January 2022, 300 million people have been infected worldwide, and 40 million people remain uncured. During this crisis, on one hand, to suppress the spread of the disease and reduce the risk of being infected, individuals spontaneously took many protective measures, such as wearing masks, disinfecting surfaces, and observing home isolation; on the other hand, many governments imposed mandatory measures to control disease spread, such as lockdowns and travel restrictions. Therefore, in the process of disease spread, the methods for inhibiting disease transmission can be mainly divided into two categories: individual spontaneous protective behaviors and mandatory government control policies. The modeling and control of infectious diseases have always been important research problems. In this paper, we provide a detailed review of two categories in the literature:

- Literature that models the spontaneous protective behaviors of individuals during disease transmission and studies how these protective behaviors affect disease transmission.
- Literature that studies how policymakers can impose mandatory measures to control the spread of disease.

The former and latter categories are for controlling the spread of disease through spontaneous protective behavior and mandatory measures, respectively. These two types of approaches encompass almost every way of suppressing the spread of the disease during the COVID-19 crisis. For the former approach, the literature mostly focuses on the modeling of individual protective behavior during the epidemic, how protective behavior affects disease transmission, and how disease transmission affects protective behavior. For the latter approach, the literature can be divided into two categories: macro- and micro-based control methods. Macro-based control methods mostly study regional policies, such as the isolation ratio of each region, the number of people allowed to move between regions, and whether a city should be closed. Micro-based control methods are studied at the network level. In these studies, the contact relationships of individuals are modeled as networks, where nodes represent individuals and edges represent contact relationships. The disease spreads in the network. These works study how to control the spread by changing the network structure, such as which nodes should be removed and which edges should be disconnected.

The rest of the paper is organized as follows. In Section 1, we introduce basic information on the disease spread model. In Section 2, we review the papers on the spontaneous protective behavior in disease spread. In Section 3, the works of mandatory measures to control the disease spread are reviewed. The conclusion and discussion are summarized in Section 4.

1 Preliminary

In this section, we briefly introduce the disease spread model, which is the basis for the studies on protective behavior analysis and spread control. The simplest disease spread model is the susceptible-infected-susceptible (SIS) model[1]. At each moment, the susceptible individual is infected with probability β by the infected individual, and the infected individual recovers with probability γ. Let $s(t)$ and $i(t)$ represent the fraction of susceptible agents and infected agents at time t, respectively, then we have

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= \gamma i(t) - s(t)i(t)\beta, \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= s(t)i(t)\beta - \gamma i(t) \end{aligned} \tag{1}$$

Eventually the proportion of infected people will reach a steady-state, i.e., homeostasis. If $\frac{\beta}{\gamma} < 1$, then $i = 1 - \frac{\gamma}{\beta}$ when it is stable, and if $\frac{\beta}{\gamma} \geqslant 1$, then $i = 0$ when it is stable. Thus 1 is the epidemic threshold of the SIS model, when the basic reproduction number

1 Department of Automation, Tsinghua University, Beijing 100084, China
Address correspondence to Wengxiang Dong, dongwx19@mails.tsinghua.edu.cn

is larger than the epidemic threshold, the disease would outbreak, otherwise, it would die out. The basic SIS model does not consider the network structure, that is, does not consider the connection relationship between individuals. When considering the network structure, the epidemic thresholds of the SIS model are different under different networks. In a homogeneous or ER network, the epidemic threshold is $\frac{1}{\langle k\rangle}$, where $\langle k\rangle$ is the average degree of the network[2]. And in the scale-free network, the epidemic threshold becomes $\frac{\langle k\rangle}{\langle k^2\rangle}$, where $\langle k^2\rangle = \sum k^2 p(k)$ is the average k^2 of each node[3]. By analyzing the steady-state and epidemic threshold, we can study the spread and control of disease in different situations.

The SIS model only considers the two states of susceptible and infected. However, for many diseases, the individual will not be reinfected after recovery. Compared with the SIS model, the susceptible-infected-recovery (SIR) model adds the recovery state[4]. Let $r(t)$ represent the fraction of recovery agents at t, then we have

$$\begin{aligned}\frac{\mathrm{d}s}{\mathrm{d}t} &= -s(t)i(t)\beta,\\ \frac{\mathrm{d}i}{\mathrm{d}t} &= s(t)i(t)\beta - \gamma i(t),\\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \gamma i(t)\end{aligned} \tag{2}$$

The SIR model does not end up in homeostasis like the SIS model because of the recovery state. Eventually, the number of infected people will be zero.

In addition, the susceptible-exposed-infected-recovery (SEIR) model adds the exposed state. Individuals in the exposed state do not show symptoms of infection but can infect others. Moreover, adding different states results in many other variants of the infectious disease model, which will not be discussed in detail here.

The goal of disease control is to minimize the number of infected people. Therefore, based on the disease spread model, the algorithm for disease control problems aims to minimize the number of infections in the disease spread model. For example, for the SIS model, the goal is to reduce the number of infections (be in state I) at equilibrium. For the SEIR model, the goal is to minimize the number of people infected (be in the states E and I) over all time periods.

2 Spontaneous Protective Behavior in Disease Spread

In the process of disease spread, the spontaneous protective behavior of individuals is crucial for inhibiting spread. On one hand, the spread of disease promotes individual protective behavior; on the other hand, individual protective behavior inhibits the spread of disease. In this section, we divide these studies into two categories: environment-induced and information dissemination-induced protective behaviors. The former category models an individual's behavior as being directly affected by the surrounding environment; the latter category believes that individual behavior is affected by the dissemination of information.

In these works, the authors mainly discussed the epidemic threshold λ under different situations. The epidemic threshold is an important indicator for measuring whether a disease can break out. When $\frac{\beta}{\gamma} > \lambda$, the disease would spread out; otherwise, it would die out. Through disease thresholds, we can analyze the impact of different protective behaviors on disease spread.

2.1 Environment-induced protective behavior

Individual behavior is largely influenced by the severity of the disease spread. Environment-induced protective behaviors are those in which individuals observe protective behaviors based on the spread in the surrounding environment. Individuals are more inclined to adopt protective behaviors if the infection rate in the environment is high.

Protective behavior can be modeled in various forms. For example, in Ref. [5], protective behaviors can change the infection rate of individual, that is

$$\beta_i = (1-\alpha)^{s_i}\beta_0 \tag{3}$$

where β_i is the actual infection rate of node i, β_0 is the initial infection rate, s_i is the number of infected neighbors of node i, and $\alpha \in [0,1]$ is a parameter that measures the strength of an individual's response.

The higher the proportion of neighbors who are infected, the stronger the individual's protective behavior and the lower the infection rate. Moreover, the larger the value of α is, the stronger the individual's response to the environment. If $\alpha = 0$, then the infection rate would always be β_0, which means that no matter how the environment changes, the individual does not practice protective behavior; if $\alpha = 1$, then if and only if there is no one around the individual infected, the infection rate is β_0, otherwise, the individual will take protective actions, and the infection rate is 0, which means that isolation or other actions are taken to completely block the spread.

On the basis of this assumption, Zhang et al.[5] obtained the epidemic thresholds for the SIS and SIR models:

$$\begin{aligned}\lambda^{SIS} &= \frac{1}{1-\alpha}\frac{\langle k\rangle}{\langle k^2\rangle} = \frac{1}{1-\alpha}\lambda_0^{SIS},\\ \lambda^{SIR} &= \frac{1}{1-\alpha}\frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle} = \frac{1}{1-\alpha}\lambda_0^{SIR}\end{aligned} \tag{4}$$

where λ_0^{SIS} and λ_0^{SIR} are the epidemic threshold for the classical SIS and SIR models, respectively. When $\frac{\beta_0}{\gamma} > \lambda_0^{SIS}$ or λ_0^{SIR}, the disease will spread out, otherwise it will die out.

After considering the influence of the environment on the individual, the epidemic threshold has become $\frac{1}{1-\alpha}$ of the original value. The epidemic threshold is only related to the strength of individual's response to the environment.

Besides, Wu et al.[6] considered the impact of individual information, local information, and global information on protective behavior. Then the infection rate is

$$\beta_i = \psi(k_i)\left(1-\alpha_2\frac{s_i}{k_i}\right)(1-\zeta p)\beta_0 \tag{5}$$

where k_i is the degree of node i and $\psi(k_i)$ is the individual awareness. The larger the node degree is, the more people the individual contacts, and the smaller $\psi(k_i)$ is, that is, the individual is more inclined to take protective behavior. α_2 is the parameter of local awareness intensity, the larger the parameter, the stronger the local awareness. ζ is the parameter of global awareness intensity, the larger the parameter, the stronger the global awareness. s_i is the number of infected neighbors of node i and $1-\alpha_2\frac{s_i}{k_i}$ is the local awareness. The higher the proportion of infection is around the individual, the stronger the protective

behavior. p is the infection density of the entire society and $1-\zeta p$ is the global awareness. The higher the infection density of the entire society, the stronger the protective behavior. Then the epidemic threshold becomes

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \psi(k) \rangle - \alpha_2 \langle k \psi(k) \rangle} \tag{6}$$

Therefore, the threshold of disease spread is affected by individual and local information but not by the impact of global information on protective behavior.

Similarly, protective behavior can be described in other ways. For instance, in Ref. [7], the infection rate was $\beta_i = e^{-(H+J(\frac{k}{k_i})^{\alpha_3})}\beta_0$. The work in Ref. [8] extended the model of Ref. [6], and modified Eq. (5) as $\beta_i = \psi(k_i)\left(1-\alpha_4\left(\frac{s_i}{k_i}\right)^{a_1}\right)(1-\zeta_2 p^{\alpha_2})\beta_0$. α_3 is the parameter of the individual's response strength. α_4 is the parameter of local awareness intensity. And ζ_2 is the parameter of global awareness intensity. The results of these studies are similar and will not be repeated here.

2.2 Information dissemination-induced protective behavior

Individual protective behavior is affected not only by the environment but also by the dissemination of information. As shown by the COVID-19 crisis, publicity about the severity of the disease largely affects the protective behavior of individuals, which in turn affects the extent of the disease spread. Therefore, some studies modeled information dissemination in the process of disease spread to study the effect of protective behavior on the disease under information dissemination.

Funk et al.[9] modeled the spread of information using an approach similar to the disease spread model. In their paper, information was modeled in three parts: information generation, information transmission, and information fading. Infected individuals generate information and spread it to others. Those who receive the information will take protective measures. In the process of dissemination, information will fade, and the stronger the information, the stronger the protective behavior. The infection rate is

$$\beta = (1-\rho^i)\beta_0 \tag{7}$$

where i represents that the information has passed through i individuals (awareness at level i) before arriving at a given individual. ρ is the decay constant, which determines how much the tendency to take the protective behavior decreases with decreasing quality of information. Therefore, the fewer people the information has passed through, the stronger the information and protective behavior, and the lower the infection rate.

The specifics of three parts of information modeling are as follows:

- **Information generation:** For the infected individual, it produces information with probability ω, for this information, the awareness level is 0.
- **Information transmission:** For the individual with information of level i, it can spread the information to another individual to level $i+1$ with probability α.
- **Information fading:** For the individual with information of level i, the information can decrease to level $i+1$ with probability λ.

Let N_i denote the population at awareness level i. Then we have

$$\frac{dN_i}{dt} = -\alpha_5 \frac{N_i}{N}\sum_0^{i-1} N_j + \alpha_5 \frac{N_{i-1}}{N}\left(N - \sum_0^{i-1} N_j\right) - \mu_2 (N_i - N_{i-1}) \tag{8}$$

where α_5 is the rate of information transmission, μ_2 is the rate of information fading, and N is the total number of individuals.

The process of disease spread is modeled by the classical SIR model. After a series of derivations, the paper discusses the basic propagation number $R(\alpha_5, \omega, \rho)$ of this model. Since the full expression of R is too complex to get a simple interpretation, they obtain its upper bounds. If the basic reproduction number is greater than

$$\lim_{\omega,\alpha_5 \to \infty} R = \frac{1}{1-\rho(1-D_k^{-1})} \tag{9}$$

the information spread cannot stop the disease from growing into an epidemic, no matter how fast the information generates and spreads.

In addition, some studies divided information spread and disease spread into two networks, establishing a dual network model[10–13]. The basic model is shown in Fig. 1 a (from Ref. [10]). The lower layer is a physical contact network, which reflects the physical contact relationships between individuals. The disease is spread by the physical contact network. The upper layer is a virtual network, which reflects the communicative relationship between individuals. The information about an epidemic is spread by the virtual network.

In Ref. [10], the disease spread process is modeled by the SIS model. The individual can be in the state of susceptible (S) or infected (I) state. The susceptible individual would be infected with a certain probability, and the infected individual would be recovered with probability γ. And for the information spread, the individual can be in the state of unaware (U) or aware (A) state. The aware individuals will take the protective measure and the

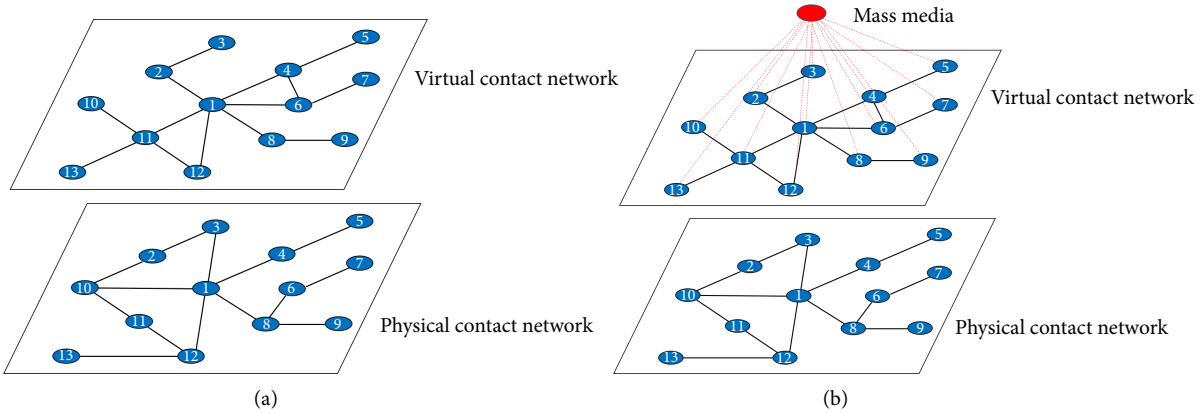

Fig. 1 (a) Two network model (from Ref. [10]). (b) Two network model with mass media (from Ref. [11]).

unaware individuals will not. Then the infection rate is β_0 for unaware individuals and $\alpha_6\beta_0$ for aware individuals, where α_6 is the parameter of the individual's awareness strength. The spread of the information is similar to that of the SIS model. The unaware individual would become aware by aware neighbor with probability η, and the aware individual would become unaware with probability ε. Let a_{ij} and b_{ij} denote the adjacency matrix of virtual contact network and physical contact network, respectively. Then the epidemic threshold is

$$\lambda_c = \frac{\beta_0}{\gamma} = \frac{1}{\Lambda_{\max}(\boldsymbol{H})} \tag{10}$$

where $\Lambda_{\max}(\boldsymbol{H})$ is the largest eigenvalue of matrix $\boldsymbol{H}$, and $h_{ji} = (1-(1-\alpha_6)p_i^A)b_{ji}$. If $\eta < \frac{\varepsilon}{\Lambda_{\max}(\boldsymbol{A})}$, then the disease spread is independent of information spread. Additionally, if $\eta > \frac{\varepsilon}{\Lambda_{\max}(\boldsymbol{A})}$, then the disease spread depends on the structure of the virtual network and the dynamic of information spread. We call $\frac{\varepsilon}{\Lambda_{\max}(\boldsymbol{A})}$ the meta-critical point. Information dissemination has an impact on the threshold of disease transmission only if the intensity of information dissemination is higher than the meta-critical point.

In addition, some studies have extended this model. For example, the model only considers information dissemination between individuals; however, a user's consciousness is largely influenced by institutions and individuals (such as the government and celebrities). In Ref. [11], a node called mass media can be added to the basic model. Every unaware individual can become aware by the mass media with probability m, which is shown in Fig. 1b (from Ref. [11]). A series of derivations can show that the influence of mass media will greatly affect the epidemic threshold and make the meta-critical point vanish. That is, the influence of mass media can suppress the spread of the epidemic.

3 Mandatory Measures to Control the Disease Spread

In addition to spontaneous protective actions, agencies such as governments can take mandatory measures to control the spread of a disease. Many researchers studied how to control the spread of disease, and we divide these studies into two types: macro-based and micro-based control methods. The former method neglects the network structure but makes macro adjustments to all individuals, such as immunizing certain nodes in proportion; the latter method is based on the network, adjusting at the micro-level, such as specifying which nodes should be immunized. Micro-based control methods are more diverse and can handle more complex problems. The method of removing nodes has a unified problem form, so in this paper, we compare the effects of different methods.

3.1 Macro-based control methods

Macro-based control methods neglect the specific network structure or consider the network structure, but the nodes on the network represent groups rather than individuals, such as communities and cities. Thus, this type of research discusses group-level decision-making.

Song et al.[14] studied how to suppress the disease spread by controlling the inter-regional mobility. The algorithm is called the dual-objective reinforcement-learning epidemic control agent (DURLECA), which adopts a GNN to capture the graph feature and uses reinforcement learning to decide on the mobility restriction between regions. Because restricting population movement can cause economic losses, the algorithm hopes to curb the spread of disease while minimizing economic losses. Then, through simulation, the authors verified that DURLECA has a better control effect and lower economic losses compared with the basic methods, such as lockdowns and restricting flow proportionally.

Wan et al.[15] proposed a multi-objective reinforcement learning framework to suppress disease spread while minimizing the cost. The framework uses regions as decision-making agents and the number of people in S, I, and R as states, establishes an optimization problem, and makes decisions by reinforcement learning. The authors verified the effectiveness of the algorithm using the data of six China cities in 2020.

Hota et al.[16] proposed a closed-loop framework that combines inferences from testing data, learning the parameters of the dynamics and optimal resource allocation for controlling the spread of disease. In this framework, each node represents a region. In the experiment, the authors considered a network with five nodes, where each node represents a country. Through simulation and the real data of COVID-19, the authors verified the effectiveness of the algorithm and found that early testing was crucial for disease control.

Birge et al.[17] studied how to suppress disease spread and minimize economic losses by the targeted closure of a city. The authors modeled disease control in each region as an optimization problem that guarantees a progressive reduction in the proportion of disease infections while minimizing economic losses. The optimization problem is

$$\begin{aligned} &\max_{x_i\in[0,1]} \sum_i c_i x_i, \\ &\text{s.t.} \quad E_i' + I_i^{c\prime} + I_i^{sc\prime} \leqslant 0 \end{aligned} \tag{11}$$

where x_i is the level of permitted economic activity in neighborhood i and c_i is a fixed location-specific constant. $E_i' + I_i^{c\prime} + I_i^{sc\prime} \leqslant 0$ means the number of infected people (E and I) declines over time.

Birge et al.[17] conducted a simulation experiment using mobile data in New York City. According to the results of the experiment, the model can restore 23.1%–42.4% of the baseline employment level on the basis of reducing infection.

Moreover, Refs. [18–20] also studied macro-based control methods. These studies simulated disease control by adjusting parameters in the disease spread model and tried to find optimal control strategies.

3.2 Micro-based control methods

(1) Preliminary: The micro-based control method is used for control at the network level. It can be divided into three tasks: remove nodes, disconnect edges, and change disease-related parameters. Removing a node means cutting off all connections between the removed node and other nodes so that it cannot spread the disease, which corresponds to isolation or immunity. Cutting edges is cutting off the connection between certain nodes, which corresponds to restricting travel. Changing disease-related parameters is to change the infection rate and recovery rate of certain individuals, which corresponds to allocating medical resources to certain individuals, etc.

Before introducing these three kinds of control methods in detail, we first introduce a conclusion, which is the theoretical basis of many micro-based control methods. In Ref. [21], Wang et al. gave the epidemic thresholds of the general network, which is shown in Theorem 1.

Theorem 1: When a disease spread in a network, then the

epidemic threshold would be

$$\lambda_c = \frac{1}{\Lambda_{\max}(\boldsymbol{A})} \tag{12}$$

where $\boldsymbol{A}$ is the adjacency matrix of the network, and $\Lambda_{\max}(\boldsymbol{A})$ is the largest eigenvalue of $\boldsymbol{A}$.

Then in Ref. [21], Wang et al. also showed how the probability of infection would change over time, which is shown in Theorem 2.

Theorem 2: When $\frac{\beta}{\gamma}$ is lower than the epidemic threshold $\left(\frac{1}{\Lambda_{\max}(\boldsymbol{A})}\right)$, the probability of infection would exponentially decrease over time.

From Theorems 1 and 2, for a given network A, if $\frac{1}{\Lambda_{\max}(\boldsymbol{A})}$ is greater than $\frac{\beta}{\gamma}$, the disease will die out and the probability of infection would exponentially decrease over time. Then for the method of removing nodes or cutting edges, if the maximum eigenvalue of the adjacency matrix is minimized, then the disease spread can be controlled.

Next we introduce three kinds of micro-based control methods in detail.

(2) Remove nodes: Given a network structure, if k nodes are to be removed, this type of research attempts to answer how these nodes should be selected to control the spread of the disease as much as possible.

The most intuitive idea is to sort the nodes according to their importance and remove the more important nodes. In Ref. [22], the authors removed nodes based on the degree. Degree is the most basic indicator to measure the connectivity of a node. In this paper, Piccini et al.[22] mainly compared the effects of the two methods: HighDegree and LowDegree. HighDegree is to remove the nodes with the highest degree, and LowDegree is to remove the nodes with the lowest degree. Piccini et al.[22] found that, in some cases, removing low-degree nodes counterintuitively had a better effect on suppressing disease spread.

Cohen et al.[23] proposed a method for removing the nodes by finding the acquaintances. In their approach, they chose a fraction p of N nodes and found a random acquaintance of theirs (a random node that connects to the chosen nodes). Then the acquaintance nodes are be removed. That means, for a node with k contacts, the probability it is selected for removal is $\frac{kP(k)}{N\langle k\rangle}$. Finally, the authors proved the effectiveness of the method through theory and experiment.

Nagaraja[24] proposed a method for suppressing disease spread by removing superspreaders. The author searched for superspreaders through random walks, and the results are similar to removal based on node degree.

Moreover, in Refs. [25, 26], the authors found the important nodes using the community structure of the network. When the network can be divided into modules, the importance of nodes between modules may be higher, and how to measure the importance of nodes in a community-based network is crucial. In Refs. [25, 26], the authors proposed mod centrality and comm centrality to measure the importance of nodes and remove nodes accordingly, respectively.

In addition to the method for removal according to node importance, from Theorem 1, we know that disease spread can be suppressed by minimizing the largest eigenvalue of the adjacency matrix. How is the change in the largest eigenvalue measured after removing nodes?

Tong et al.[27] proposed a method called Netshield, which measures the change in the maximum eigenvalue of the matrix after removing the nodes using the shield value. The shield-value of a network is defined as

$$Sv(S) = \sum_{i\in S} 2\lambda_2 \boldsymbol{u}(i)^2 - \sum_{i,j\in S} \boldsymbol{A}(i,j)\boldsymbol{u}(i)\boldsymbol{u}(j) \tag{13}$$

where S is the node set, $\boldsymbol{A}$ is the adjacency matrix, λ_2 is the largest eigenvalue of $\boldsymbol{A}$, and $\boldsymbol{u}$ is the largest eigenvector of $\boldsymbol{A}$.

In Ref. [28], the authors gave an approximate proof. Let $\lambda^{(S)}$ be the largest eigenvalue of the adjacency matrix after removing S from the original point set. Then,

$$\lambda_2 - \lambda^{(S)} = Sv(S) + O\left(\sum_{j\in S} \|\boldsymbol{A}(:,j)\|^2\right) \tag{14}$$

That is, the shield value approximates the largest eigenvalue change. Then, by maximizing the shield value, we can approximately minimize the largest eigenvalue. The algorithm would find a near-optimal set of points to reduce the largest eigenvalue to control the disease spread.

Then in Ref. [28], the authors improved the Netshield algorithm and proposed the Netshield+ algorithm, which reduces the time complexity of the algorithm.

In addition to Netshield, Refs. [29, 30] also adopted a similar idea to suppress disease spread by minimizing the largest eigenvalue. In Ref. [29], Ahmad et al. proposed a method called GreedyDrop that minimizes the largest eigenvalue by a spectral method. In Ref. [30], Saxena et al. used the method Shapley value based information delimiters (SVID) based on the group-based game theoretic payoff division approach to minimize the largest eigenvalue.

Considering that many control methods are available for removing nodes, and they deal with the same problems, in this paper, we compare the effects of different methods. These methods are: Netshield[27], Netshield+[28], Acquaintance[23], GreedyDrop[29], HighDegree[22], LowDegree[22], SuperSpreader[24], SVID[30], ModCentrality[25], and CommCentrality[26]. We conduct experiments on a scale-free network of 1000 nodes, and the average degree of the network is 6, each experiment is repeated 100 times and averaged, the result is shown in Fig. 2.

Figure 2a is the result with different λ ($\lambda = \frac{\beta}{\gamma}$, and the number of removed nodes is fixed as 100). The infection rate represents the probability of infection when the disease is stable. Figure 2b is the result with different number of removed nodes (λ is fixed as 1). Generally, among the methods based on minimizing the largest eigenvalue, GreedyDrop has the best effect, followed by Netshield+. However, the effect of these methods is not much different from that of HighDegree, which directly removes nodes according to the degree, and that of SuperSpreader, which removes nodes according to the importance of the nodes. In all cases, the effect of preferentially removing low-degree nodes (LowDegree) is the worst. Overall, these algorithms are sensitive to the base reproduction number and the number of removed nodes. When the base reproduction number is high, or the number of removed nodes is low, the effects of these algorithms are significantly reduced.

(3) Cut off edges: In addition to removing nodes, cutting edges is also an important means of inhibiting the spread of diseases. Similar to removing nodes, minimizing the adjacency matrix's largest eigenvalue by cutting edges is also an effective way to suppress disease spread. In Ref. [31], Tong et al. considered both removing edges to minimize spread and adding edges to

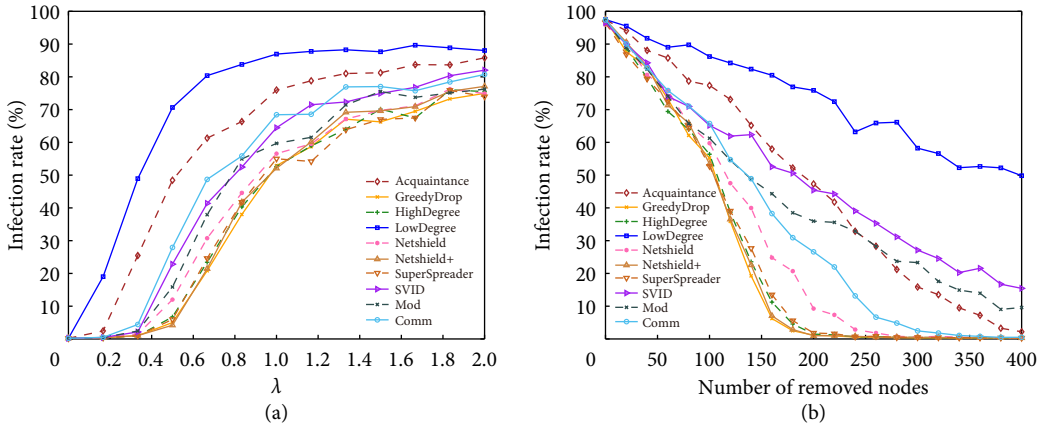

Fig. 2 (a) Comparition of algorithms with different λ. (b) Comparition of algorithms with different numbers of removed nodes.

maximize spread. Here, we focus on the method to suppress the propagation. Similar to Netshield, the authors also proposed an indicator to reflect the amount of change in the largest eigenvalue after removing edges. The algorithm is called K-EDGEDELETION, which minimizes the largest eigenvalue by removing edges to control disease spread.

In addition to K-EDGEDELETION, Refs. [32, 33] also studied how to suppress the spread by cutting off edges. The idea is still to minimize the largest eigenvalue of the matrix, we will not introduce it in detail here.

(4) Change disease-related parameters: In the epidemic model, many parameters, including the infection rate and the recovery rate, can be changed. By changing these parameters, the spread can be controlled. However, changing the parameters has a cost, and most of these papers studied how to change the parameters under a limited budget to control the spread as much as possible. Moreover, the infection rate and recovery rate of individuals can also be changed by allocating medical resources. Therefore, this type of research explored how to allocate resources to control the spread of a disease to the greatest extent.

In Ref. [34], Gourdin et al. studied how to find the best curing strategy to minimize the steady-state infection vector, $\boldsymbol{V}_{\text{inf}} = [v_1, v_2, \ldots, v_N]^{\text{T}}$, where v_i denotes the probability that node i is infected at steady-state. The problem is formulated as

$$\begin{aligned} &\min_{\Delta} f_1(\Delta) = \sum_i v_i(\Delta), \\ &\text{s.t.} \quad \sum_i \delta_i = 2L\alpha_7\beta, \\ &0 \leqslant \delta_i \leqslant \delta_c \end{aligned} \tag{15}$$

where L is the number of edges in the network, β is the infection rate, and α_7 is a parameter in $[0,1]$. That is, the sum of the recovery rates of all individuals has a constraint, which is related to the number of network connections and the infection rate, and the recovery rate of each individual has an upper limit. On this basis, Gourdin et al.[34] tried to minimize the overall infection probability when it is stable. Afterward, the authors transformed the problem into a convex optimization problem through optimization theory and verify its effectiveness through simple experiments.

Preciado et al.[35] considered the allocation of multiple types of conservation resources: preventive resources able to prevent the individual from being infected and corrective resources able to promote individual recovery after infection. Similar to Theorem 2, when the real part of $\lambda_1(diag(\beta_i)A_G - diag(\delta_i))$ is lower than a certain value, where λ_1 is the largest eigenvalue of the adjacency matrix, the probability of infection would exponentially decline over time. Therefore, the goal of the paper is to minimize the cost on the basis of ensuring that the largest eigenvalue is below a certain value. So the optimization problem is

$$\begin{aligned} &\min_{\delta_i,\beta_i} \sum_i f_i(\delta_i) + g_i(\beta_i), \\ &\text{s.t.} \quad \mathcal{R}[\lambda_1(diag(\beta_i)A_G - diag(\delta_i))] \leqslant -\varepsilon, \\ &\underline{\beta_i} \leqslant \beta_i \leqslant \overline{\beta_i}, \\ &\underline{\delta_i} \leqslant \delta_i \leqslant \overline{\delta_i} \end{aligned} \tag{16}$$

where $\mathcal{R}[\lambda_1(diag(\beta_i)A_G - diag(\delta_i))]$ is the real part of $\lambda_1(diag(\beta_i)A_G - diag(\delta_i))$. $\underline{\beta_i}$ and $\underline{\delta_i}$ are the lower bounds of β_i and δ_i. $\overline{\beta_i}$ and $\overline{\delta_i}$ are the upper bounds of β_i and δ_i.

In Ref. [36, 37], the authors studied how to allocate the resources for a competitive epidemic. In a competitive epidemic, an individual can be infected by two diseases but not simultaneously. Similarly, this type of problem can also be modeled as an optimization problem, which will not be introduced in detail here.

4 Conclusion and Discussion

In this paper, we conduct a detailed review of two categories in the literature. For the literature on an individual's spontaneous protective behaviors, the models can be divided into two types: environment-induced and information-induced. These two types of studies model from different perspectives and derive conclusions that are consistent with reality. In the literature on mandatory measures for disease control, the measures can be divided into two types: macro-based and micro-based control methods. The former method studies the area-based control method, whereas the latter method considers the network structure, and its control method is based on the individual. Micro-based control methods can be divided into three categories: removing nodes, cutting edges, and changing disease-related parameters. Among the node removal methods, the control method by minimizing the largest eigenvalue of the adjacency matrix has the best effect, but the difference with the most basic method of removing nodes according to the degree of the node is not obvious. The deficiencies of the modeling in the literature are as follows. For the literature on individual spontaneous protective behaviors, existing methods do not consider some more realistic scenarios, such as the spread of false information and the effect of individual panic. For the literature on mandatory measures of disease control, existing methods still need to be improved in efficiency and effectiveness. Because disease transmission involves large-scale populations, individual behavior modeling and transmission control in disease transmission are important

research issues in mass intelligence.

Dates

Received: 23 February 2022; Revised: 22 March 2022; Accepted: 23 March 2022

References

[1] A. Lajmanovich and J. A. Yorke, A deterministic model for gonorrhea in a nonhomogeneous population, *Mathematical Biosciences*, vol. 28, nos. 3&4, pp. 221–236, 1976.

[2] J. O. Kephart and S. R. White, Directed-graph epidemiological models of computer viruses, in *Computation: The Micro and the Macro View*, B. A. Huberman, ed. Singapore: World Scientific, 1992, pp. 71–102.

[3] R. Pastor-Satorras and A. Vespignani, Epidemic spreading in scale-free networks, *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

[4] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927.

[5] H. -F. Zhang, J. -R. Xie, M. Tang, and Y. -C. Lai, Suppression of epidemic spreading in complex networks by local information based behavioral responses, *Chaos*: *An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 4, p. 043106, 2014.

[6] Q. Wu, X. Fu, M. Small, and X. -J. Xu, The impact of awareness on epidemic spreading in networks, *Chaos*: *An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 1, p. 013101, 2012.

[7] F. Bagnoli, P. Liò, and L. Sguanci, Risk perception in epidemic modeling, *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 6, p. 061904, 2007.

[8] Y. Shang, Modeling epidemic spread with awareness and heterogeneous transmission rates in networks, *Journal of Biological Physics*, vol. 39, no. 3, pp. 489–500, 2013.

[9] S. Funk, E. Gilad, C. Watkins, and V. A. Jansen, The spread of awareness and its impact on epidemic outbreaks, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6872–6877, 2009.

[10] C. Granell, S. Gómez, and A. Arenas, Dynamical interplay between awareness and epidemic spreading in multiplex networks, *Physical Review Letters*, vol. 111, no. 12, p. 128701, 2013.

[11] C. Granell, S. Gómez, and A. Arenas, Competing spreading processes on multiplex networks: Awareness and epidemics, *Physical Review E*, vol. 90, no. 1, p. 012808, 2014.

[12] Z. Wang, Q. Guo, S. Sun, and C. Xia, The impact of awareness diffusion on SIR-like epidemics in multiplex networks, *Applied Mathematics and Computation*, vol. 349, pp. 134–147, 2019.

[13] C. Zheng, C. Xia, Q. Guo, and M. Dehmer, Interplay between SIR-based disease spreading and awareness diffusion on multiplex networks, *Journal of Parallel and Distributed Computing*, vol. 115, pp. 20–28, 2018.

[14] S. Song, Z. Zong, Y. Li, X. Liu, and Y. Yu, Reinforced epidemic control: Saving both lives and economy, arXiv preprint arXiv: 2008.01257, 2020.

[15] R. Wan, X. Zhang, and R. Song, Multi-objective reinforcement learning for infectious disease control with application to COVID-19 spread, arXiv preprint arXiv: 2009.04607, 2020.

[16] A. R. Hota, J. Godbole, and P. E. Pare, A closed-loop framework for inference, prediction, and control of SIR epidemics on networks, *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2262–2278, 2021.

[17] J. R. Birge, O. Candogan, and Y. Feng, Controlling epidemic spread: Reducing economic losses with targeted closures, http://dx.doi.org/10.2139/ssrn.3590621, 2020.

[18] T. Andersson, A. Erlanson, D. Spiro, and R. Östling, Optimal trade-off between economic activity and health during an epidemic, arXiv preprint arXiv: 2005.07590, 2020.

[19] D. W. Berger, K. F. Herkenhoff, and S. Mongey, An SEIR infectious disease model with testing and conditional quarantine, Tech. Rep. 597, National Bureau of Economic Research, New York, NY, USA, 2020.

[20] P. D. Fajgelbaum, A. Khandelwal, W. Kim, C. Mantovani, and E. Schaal, Optimal lockdown in a commuting network, *American Economic Review: Insights*, vol. 3, no. 4, pp. 503–522, 2021.

[21] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in *Proc. 22nd International Symposium on Reliable Distributed Systems*, Florence, Italy, 2003, pp. 25–34.

[22] J. Piccini, F. Robledo, and P. Romero, Node-immunization strategies in a stochastic epidemic model, in *Proc. First International Workshop on Machine Learning, Optimization, and Big Data*, Sicily, Italy, 2015, pp. 222–232.

[23] R. Cohen, S. Havlin, and D. Ben-Avraham, Efficient immunization strategies for computer networks and populations, *Physical Review Letters*, vol. 91, no. 24, p. 247901, 2003.

[24] S. Nagaraja, Unlinking super-linkers: The topology of epidemic response (COVID-19), arXiv preprint arXiv: 2006.02241, 2020.

[25] N. Masuda, Immunization of networks with community structure, *New Journal of Physics*, vol. 11, no. 12, p. 123018, 2009.

[26] N. Gupta, A. Singh, and H. Cherifi, Centrality measures for networks with community structure, *Physica A*: *Statistical Mechanics and its Applications*, vol. 452, pp. 46–59, 2016.

[27] H. Tong, B. A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau, On the vulnerability of large graphs, in *Proc. 2010 IEEE International Conference on Data Mining*, Sydney, Australia, 2010, pp. 1091–1096.

[28] C. Chen, H. Tong, B. A. Prakash, C. E. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau, Node immunization on large graphs: Theory and algorithms, *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 1, pp. 113–126, 2015.

[29] M. Ahmad, J. Tariq, M. Shabbir, and I. Khan, Spectral methods for immunization of large networks, arXiv preprint arXiv: 1711.00791, 2017.

[30] C. Saxena, M. Doja, and T. Ahmad, Group based centrality for immunization of complex networks, *Physica A*: *Statistical Mechanics and Its Applications*, vol. 508, pp. 35–47, 2018.

[31] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos, Gelling, and melting, large graphs by edge manipulation, in *Proc. 21st ACM International Conference on Information and Knowledge Management*, Maui, HI, USA, 2012, pp. 245–254.

[32] C. J. Kuhlman, G. Tuli, S. Swarup, M. V. Marathe, and S. S. Ravi, Blocking simple and complex contagion by edge removal, in *Proc. 2013 IEEE 13th International Conference on Data Mining*, Dallas, TX, USA, 2013, pp. 399–408.

[33] C. Chen, H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos, Eigen-optimization on large graphs by edge manipulation, *ACM Transactions on Knowledge Discovery from Data* (*TKDD*), vol. 10, no. 4, pp. 1–30, 2016.

[34] E. Gourdin, J. Omic, and P. V. Mieghem, Optimization of network protection against virus spread, in *Proc. 2011 8th International Workshop on the Design of Reliable Communication Networks* (*DRCN*), Krakow, Poland, 2011, pp. 86–93.

[35] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas, Optimal resource allocation for network protection against spreading processes, *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 99–108, 2014.

[36] N. J. Watkins, C. Nowzari, V. M. Preciado, and G. J. Pappas, Optimal resource allocation for competing epidemics over arbitrary networks, in *Proc. 2015 American Control Conference* (*ACC*), Chicago, IL, USA, 2015, pp. 1381–1386.

[37] N. J. Watkins, C. Nowzari, V. M. Preciado, and G. J. Pappas, Optimal resource allocation for competitive spreading processes on bilayer networks, *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 298–307, 2016.