PAPER • OPEN ACCESS

Efficient network immunization strategy based on generalized Herfindahl–Hirschman index

To cite this article: Peng Chen *et al* 2021 *New J. Phys.* **23** 063064

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
24 March 2021

REVISED
12 May 2021

ACCEPTED FOR PUBLICATION
27 May 2021

PUBLISHED
23 June 2021

PAPER

Efficient network immunization strategy based on generalized Herfindahl–Hirschman index

Peng Chen[1,6], **Mingze Qi**[1,6], **Xin Lu**[2,6], **Xiaojun Duan**[1,*] **and Jürgen Kurths**[3,4,5]

[1] College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, Hunan, 410073, People's Republic of China
[2] College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, 410073, People's Republic of China
[3] Department of Physics, Humboldt University, Newtonstr. 15, Berlin, 12489, Germany
[4] Potsdam Institute for Climate Impact Research, P.O. Box 60 12 03, Potsdam, 14412, Germany
[5] Centre for Analysis of Complex Systems, World-Class Research Center 'Digital biodesign and personalized healthcare', Sechenov First Moscow State Medical University, Moscow, 119991, Russia
[*] Author to whom any correspondence should be addressed.
[6] These authors contributed equally to this work.

E-mail: xjduan@nudt.edu.cn

Keywords: network immunization, disease spreading, generalized Herfindahl–Hirschman index, vaccination strategy

Abstract

The topic of finding effective strategies to restrain epidemic spreading in complex networks is of current interest. A widely used approach for epidemic containment is the fragmentation of the contact networks through immunization. However, due to the limitation of immune resources, we cannot always fragment the contact network completely. In this study, based on the size distribution of connected components for the network, we designed a risk indicator of epidemic outbreaks, the generalized Herfindahl–Hirschman index (GHI), which measures the upper bound of the expected infection's prevalence (the fraction of infected nodes) in random outbreaks. An immunization approach based on minimizing GHI is developed to reduce the infection risk for individuals in the network. Experimental results show that our immunization strategy could effectively decrease the infection's prevalence as compared to other existing strategies, especially against infectious diseases with higher infection rates or lower recovery rates. The findings provide an efficient and practicable strategy for immunization against epidemic diseases.

1. Introduction

The spreading phenomenon is a pervasive process in nature that describes many essential activities in society, such as infectious disease outbreaks, information dissemination, viral marketing, etc [1–4]. Nowadays, a potential pandemic can possibly reach every city in the world within a few days, which allows for a local disease to evolve into a global pandemic. This is what happened with COVID-19 [5–8]. It is thus urgent and essential to design efficient mechanisms for the restraint of epidemic spreading.

Complex networks have been proven to be a powerful analytical tool for predicting and controlling epidemic spreading in real-world scenarios [9–11]. These infectious diseases are transmitted in a population through the network of contacts between individuals. One of the critical problems is how to best distribute limited treatment and vaccination resources to suppress disease outbreaks [12, 13]. There has been an abundant production of heuristic rankings [14–17] for vaccination or quarantine to identify influential nodes in networks. Some local strategies, such as acquaintance immunization [18] and random-walk immunization [19], have been introduced when the complete knowledge of all individuals is not known. The sampling method was also considered for the immunization strategy of hidden populations [20]. Moreover, recent studies [21, 22] seeking immunization strategies have applied message passing techniques, which consider both the network topology and epidemic dynamic. In fact, the immunization problem is similar to the network disintegration problem [23–25], which focuses on the destruction of harmful networks through targeted attacks.

The network disintegration problem focuses on determining a set of vertices or links whose removal would collapse the giant connected component (GCC). The most traditional solutions to this problem are

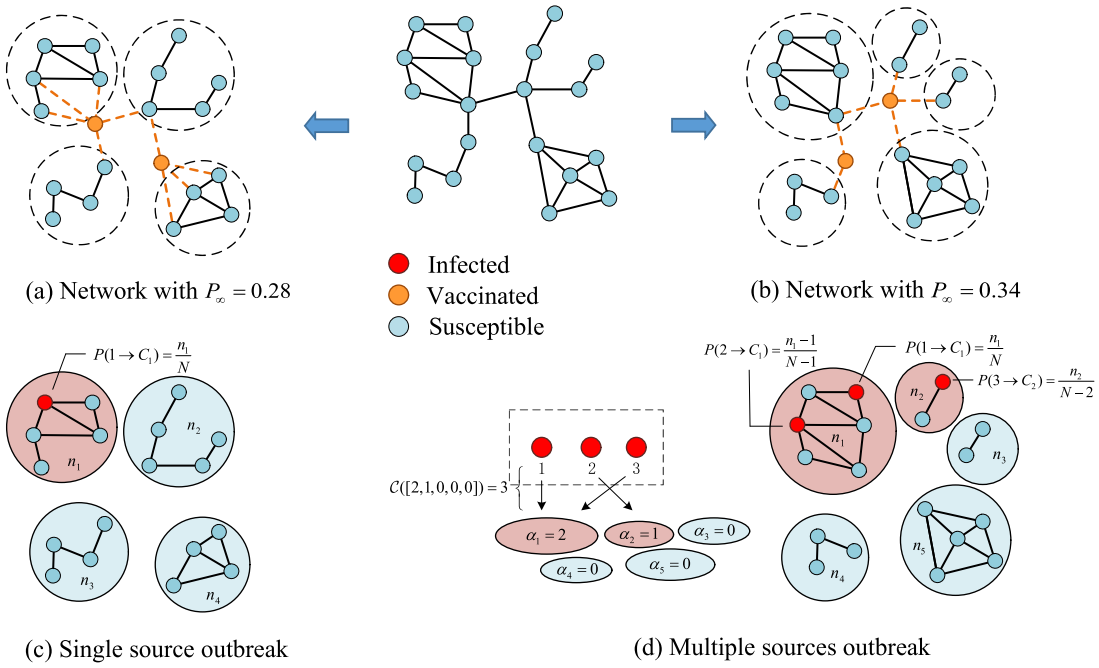

Figure 1. Schematic illustration of GHI. For two different immunization (attack) strategies, (a) one causes the remaining network with $N = 18$, $P_\infty = 0.28$, and (b) the other leads to the remaining network with $N = 18$, $P_\infty = 0.34$, and where P_∞ is the proportion of nodes in the GCC. (c) The network's case is infected by an infectious disease that hits a node (red) randomly, and the nodes in the component have the risk of being infected. The expectation of the risk region's size in random outbreaks starting from a single random infected node is the HHI. (d) The infectious disease has randomly infected three nodes (red) in the network. The expectation of the total size of the risk region in random outbreaks starting from multiple random nodes is GHI.

node ranking methods [26–28], which identify the sequence of nodes that will maximize the damage to the network's connectivity. The importance of nodes is often represented by node degree, betweenness, or k-shell centrality, etc [29]. Significantly, the adaptive centrality strategy, which recalculates the centrality of the undismantled nodes at each step, can dramatically improve the effect of dismantling [30]. Recently, several practical algorithms have been proposed for dismantling a network based on collective influence (CI) [31], decycling and tree breaking [32, 33], optimal partitioning of graphs [23, 34], explosive percolation [24], or articulation points [35]. Moreover, combinatorial optimization-based approaches, including tabu search [36, 37], evolutionary algorithm [25], and the deep reinforcement learning framework [38], have been presented to search for the optimal disintegration strategy.

Connectivity is necessary for a network to maintain its function, so the complete fragmentation of a network is the common goal of network disintegration and immunization. However, it is infeasible to fragment the network entirely in many real cases e.g., when vaccine resources are limited, or maintaining the normal operation of society requires certain liquidity under travel restrictions [39, 40]. That is to say, there are a certain size and quantity of connected components in the network after immunization. In this context, not all possible spreading occurs in the GCC, and other connected components also contribute to the spread of infectious diseases. Therefore, a new index is required to evaluate the spreading risk of infectious diseases in a network and design effective immunization strategies.

In this paper, we propose an indicator based on the size distribution of connected components to measure the connectivity of the network. Minimizing this index by removing a set of nodes or links, we could obtain an efficient immunization strategy that minimizes the infection risk of individuals in networks. The remainder of this paper is organized as follows. In section 2, we present an index named the generalized Herfindahl–Hirschman index (GHI) and a fast method to approximate the GHI is given in section 3. In section 4, the GHI-based optimization model is proposed to design an immunization strategy. The effects and characteristics of the strategy are discussed through experiments in different networks and spreading models. Finally, the conclusion and discussion are presented in section 5.

2. The definition of generalized Herfindahl–Hirschman index

Complex networks have long been acknowledged as a key ingredient of epidemic modeling [41, 42], which describes how individuals interact with one another. A complex network can be described as an undirected graph $G = (V, E)$, where V is the set of nodes, and $E \subseteq V \times V$ is the set of edges. $N = |V|$ and $M = |E|$ are

the number of nodes and edges in the network, respectively. The spreading process in the network depends on the network connectivity. The essence of immunization is to fragment the transmission network into small connected components, the largest of which is the GCC, and the GCC's size is a common network connectivity measure. However, when we evaluate the potential risk of infectious diseases, the size of the GCC cannot reflect the infection risk of other components, in which infectious diseases can also break out. Therefore, the infection risk of individuals in a network cannot be judged directly by the GCC. For example, with the two networks presented in figure 1, it is not clear whether the network in figure 1(a) has a lower risk of infection, even though it has a smaller GCC than the network in figure 1(b).

As shown in figure 1(c), the network contains four connected components after immunization. We assume that a disease hits a random node in the network, and therefore all the nodes in the connected component containing the infected node have the risk of being infected. The infection risk of the nodes in the network is defined as the expected fraction of nodes at risk of infection in random outbreaks. In an epidemic model, the nodes can be divided into different states, such as susceptible (S), infected (I), or recovered (R), while the links allow contagion between the nodes. The susceptible–infectious (SI) epidemic spreading model [1, 43] represents an infectious disease spread in which infected individuals never recover and keep propagating the disease forever. In the SI model, all the nodes of the connected component will be infected if one node of the connected component becomes infected. Therefore, the infection risk of the individual in the network is described as the expectation of an infection's prevalence (the fraction of infected nodes) in random outbreaks in the SI model, which can be approximated by simulation. In this mechanism, we deduce an accurate expression for calculating the infection risk of the nodes. Say that n_i is the size of components C_i in the network, where $i = 1, \ldots, L$ represents the serial number of components, and $p_i = n_i/N$ is the proportion of nodes in components C_i. Accordingly, the probability that a random outbreak starts in component C_i is equal to p_i, and the average number of nodes at risk of infection under the infection of a random node is

$$\langle R_{\text{risk}} \rangle = \sum_{i=1}^{L} n_i p_i = N \sum_{i=1}^{L} p_i^2. \tag{1}$$

After normalization, the expression of the infection risk of the nodes is equivalent to the Herfindahl–Hirschman index (HHI), denoted by ϕ.

$$\phi = \frac{\langle R_{\text{risk}} \rangle}{N} = \sum_{i=1}^{L} \left(\frac{n_i}{N} \right)^2. \tag{2}$$

The HHI [44] is a commonly accepted measure of market concentration in economics. It is calculated by squaring each firm's market share competing in the market and then summing the resulting numbers.

In epidemic outbreaks, the spreading usually starts from multiple infected nodes. Therefore, we generalized HHI to the GHI to measure the infection risk of the nodes from a multi-sourced infection. The distribution of the infection sources in all L connected components is denoted by $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_L)$, where $0 \leqslant \alpha_k \leqslant n_k$, and $\sum_{k=1}^{L} \alpha_k = \Omega$ is the number of the initial infection sources. Calculating the probability of $\boldsymbol{\alpha}$ can be regarded as the problem of placing different sources of infection in different parts (nodes) of the network. The problem is divided into two steps: the first step is to determine which connected component each infection source corresponds to, and then the second step is to assign Ω infection sources to different nodes in the corresponding component. For the infection source distribution $\boldsymbol{\alpha}$, it contains all the results obtained by sampling the infection sources according to the group division $(\alpha_1, \alpha_2, \ldots, \alpha_L)$, and the total number of all sampling results is

$$\mathcal{C}(\boldsymbol{\alpha}) = \frac{\Omega!}{\alpha_1! \alpha_2! \ldots \alpha_L!}. \tag{3}$$

Then, the probability of the distribution of the infection sources $\boldsymbol{\alpha}$ is

$$\mathcal{P}(\boldsymbol{\alpha}) = \mathcal{C}(\boldsymbol{\alpha}) \frac{(N-\Omega)!}{N!} \prod_{1 \leqslant i \leqslant L} \frac{n_i!}{(n_i - \alpha_i)!}. \tag{4}$$

For example, in figure 1(d), three nodes are initially infected in the network and $\boldsymbol{\alpha} = [2, 1, 0, 0, 0]$. First, different sources of infection are marked with the order 1, 2, 3 and placed into the components. The total number of placement methods is $\mathcal{C}([2, 1, 0, 0, 0]) = 3$. After determining the corresponding relationship between the infection sources and the connected components, the sources are put into the network in order. For a certain placement method such as $1 \to C_1, 2 \to C_1, 3 \to C_2$, the probability of putting the infection source 1 in C_1 is n_1/N. After node 1 is infected, subsequent nodes cannot re-infect node 1, so the

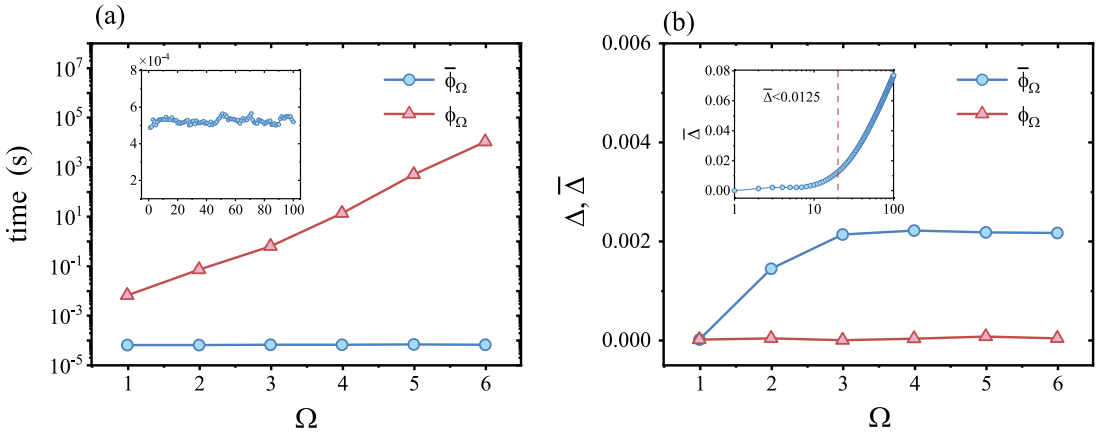

Figure 2. The time cost and accuracy of $\overline{\phi_\Omega}$ (blue line) and ϕ_Ω (red line) on Erdős–Rényi (ER) networks. (a) The relationship of the $\overline{\phi_\Omega}$, ϕ_Ω running time and the number of initial infection sources Ω. In the inset of (a), the time cost of $\overline{\phi_\Omega}$ is extended from $\Omega = 1$ to $\Omega = 100$. (b) The relationship of $\overline{\Delta}$, Δ and Ω on the ER network with $N = 100$ and average degree $\langle k\rangle = 1.5$. The $\overline{\Delta}$, Δ is the absolute value of the difference between $\overline{\phi_\Omega}$, ϕ_Ω and ρ_1 (the average infection's prevalence of outbreaks) in 10^6 simulations of the SI model. In the inset of (b), the $\overline{\Delta}$ calculated by $\overline{\phi_\Omega}$ is extended from $\Omega = 1$ to $\Omega = 100$, and $\overline{\Delta}$ is less than 0.0125 on the left side of the dashed line. The dashed line represents the case of $\Omega/N = 0.2$.

probability of putting the infection source 2 in C_1 then becomes $(n_1 - 1)/(N - 1)$. Similarly, the probability of putting the infection source 3 in C_2 becomes $n_2/(N - 2)$. The probabilities are the same for different initial infection results under the same infection source distribution. Finally, the probabilities of all possible outcomes are added together to get the probability of the distribution $\boldsymbol{\alpha}$. Considering the set of all possible $\boldsymbol{\alpha}$: $\mathcal{A} = \left\{\boldsymbol{\alpha} | \, 0 \leqslant \alpha_k \leqslant n_k, \sum_{i=1}^{L} \alpha_k = \Omega, \ k = 1, \ldots, L\right\}$, the GHI ($\phi_\Omega$) is defined as

$$\phi_\Omega = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \left[\mathcal{P}(\boldsymbol{\alpha}) \sum_{1\leqslant i\leqslant L} I(\alpha_i)\, p_i\right], \tag{5}$$

where $I(\alpha_i) = \begin{cases} 1, & \alpha_i > 0 \\ 0, & \alpha_i = 0 \end{cases}$ is the characteristic function of α_i, and $\sum_{1\leqslant i\leqslant L} I(\alpha_i)\, p_i$ is the total proportion of nodes at risk of infection (nodes belonging to infected components). The GHI is equal to Ω/N when the network consists of many small components of relatively equal size. In this case, GHI approaches 0 if the initial number of infected nodes Ω is very small compared to N ($\phi_\Omega = \Omega/N, \Omega \ll N, \phi_\Omega \to 0$). In addition, GHI reaches its maximum 1 when the network is connected ($\phi_\Omega = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \mathcal{P}(\boldsymbol{\alpha}) = 1$). The GHI is formally equivalent to the HHI when $\Omega = 1$. Using (5), the infection risk of the two networks can be calculated as $\phi_1 = 0.2531$, $\phi_3 = 0.607$ in figure 1(a) and $\phi_1 = 0.2407$, $\phi_3 = 0.5673$ in figure 1(b), respectively. The network in figure 1(a) has a greater risk of infection than the network in figure 1(b) although it has a smaller GCC.

It is notably complicated to use equation equation (5) to calculate ϕ_Ω because there are numerous combinations of $\boldsymbol{\alpha}$. The number of possible situations increases exponentially with the number of initial infections and connected components.

3. Approximation of the GHI

In this section, an approximate expression of GHI is considered for fast calculation. The actual initial number of infection sources is much smaller than the total number of individuals (i.e., $\Omega \ll N$), so we obtain

$$\frac{(N-\Omega)!}{N!} = \frac{1}{N(N-1)\ldots(N-\Omega+1)} \approx \frac{1}{N^\Omega} = \prod_{1\leqslant i\leqslant L} \frac{1}{N^{\alpha_i}} \tag{6}$$

and

$$\frac{n_i!}{(n_i-\alpha_i)!} = n_i(n_i-1)\ldots(n_i-\alpha_i+1) \approx {n_i}^{\alpha_i}, \tag{7}$$

and equation (4) is simplified to

$$\overline{\mathcal{P}}(\boldsymbol{\alpha}) = \mathcal{C}(\boldsymbol{\alpha}) \prod_{1\leqslant i\leqslant L} \left(\frac{n_i}{N}\right)^{\alpha_i} = \mathcal{C}(\boldsymbol{\alpha}) \prod_{1\leqslant i\leqslant L} p_i^{\alpha_i}. \tag{8}$$

Meanwhile, we could relax the restriction $\alpha_k \leqslant n_k$ in $\mathcal{A}$, when $\Omega \ll N$, i.e., one allows cases for which $\alpha_k > n_k$, and obtain $\overline{\mathcal{A}} = \left\{\boldsymbol{\alpha} |\, \alpha_k \geqslant 0, \sum_{i=1}^{L} \alpha_k = \Omega,\ k = 1, \ldots, L\right\}$. For set $\overline{\mathcal{A}}$, $\alpha_k \leqslant n_k$ holds for most connected components when $\Omega \ll N$. Therefore, replacing $\mathcal{A}$ with $\overline{\mathcal{A}}$ has little effect on the result of calculating GHI.

$$\begin{aligned}
\phi_\Omega &\approx \sum_{\boldsymbol{\alpha}\in\overline{\mathcal{A}}} \left(\overline{\mathcal{P}}(\boldsymbol{\alpha}) \sum_{1\leqslant i\leqslant L} \left(I(\alpha_i)\, p_i\right) \right) \\
&= \sum_{1\leqslant i\leqslant L} \left(p_i \times \sum_{\boldsymbol{\alpha}\in\overline{\mathcal{A}}\cap\{\boldsymbol{\alpha}|\alpha_i\neq 0\}} \overline{\mathcal{P}}(\boldsymbol{\alpha}) \right) \\
&= \sum_{1\leqslant i\leqslant L} p_i \left(1 - \sum_{\boldsymbol{\alpha}\in\overline{\mathcal{A}}\cap\{\boldsymbol{\alpha}|\alpha_i= 0\}} \overline{\mathcal{P}}(\boldsymbol{\alpha}) \right) \\
&= \sum_{1\leqslant i\leqslant L} p_i \left(1 - \sum_{\boldsymbol{\alpha}\in\overline{\mathcal{A}}\cap\{\alpha|\alpha_i= 0\}} \frac{\Omega!}{\alpha_1!\alpha_2!\ldots\alpha_L!} \prod_{1\leqslant j\leqslant L} p_j^{\alpha_i} \right).
\end{aligned} \tag{9}$$

For $\alpha_i = 0$, we can use the multinomial theorem to simplify equation (9) and approximate equation (5) as

$$\begin{aligned}
\overline{\phi_\Omega} &= \sum_{1\leqslant i\leqslant L} p_i \left(1 - \sum_{\boldsymbol{\alpha}\in\overline{\mathcal{A}}\cap\{\boldsymbol{\alpha}|\alpha_i= 0\}} \frac{\Omega!}{\alpha_1!\ldots\alpha_{i-1}!\alpha_{i+1}!\ldots\alpha_L!} \prod_{j\neq i, 1\leqslant j\leqslant L} p_j^{\alpha_i} \right) \\
&= \sum_{1\leqslant i\leqslant L} p_i \left(1 - \left(\sum_{j\neq i, 1\leqslant j\leqslant L} p_j \right)^{\Omega} \right) \\
&= \sum_{1\leqslant i\leqslant L} p_i \left(1 - \left(1 - p_i\right)^{\Omega} \right),
\end{aligned} \tag{10}$$

in which $1 - \left(1 - p_i\right)^{\Omega}$ is the probability of infection in the connected component C_i. Hence, the physical explanation for equation (10) is the sum of expected infection risk for each connected component. The cause of the discrepancy between equation (10) and the precise form equation (5) is that equation (10) allows different sources to infect the same node in the network repeatedly. However, the possibility of a repeated infection is small when $\Omega \ll N$. So it is reasonable to use $\overline{\phi_\Omega}$ as an approximation of ϕ_Ω. Equation (10) is also equivalent to the precise form equation (5) when $\Omega = 1$.

Next, we compare the computation times for ϕ_Ω and $\overline{\phi_\Omega}$ in the Erdős–Rényi (ER) network with different numbers of initial infected nodes. The results are shown in figure 2(a), the computation time of ϕ_Ω increases exponentially with the number of infected nodes, preventing it from being executed even with few initial infection sources. The running time can be considerably reduced by the approximate calculation $\overline{\phi_\Omega}$. Due to the high time complexity of ϕ_Ω, we use the ρ_{I} instead of ϕ_Ω to verify the effectiveness of $\overline{\phi_\Omega}$ in subsequent experiments, where ρ_{I} is the average infection's prevalence (the fraction of infected nodes) obtained from simulations of the SI model. In the simulation, we let each node is initially infected with the probability I_0, and then iterate the SI process with synchronous updating. After the system reaches the state, ρ_{I} is obtained. To verify the effectiveness of ϕ_Ω and $\overline{\phi_\Omega}$, we present a comparison between ϕ_Ω, $\overline{\phi_\Omega}$ and ρ_{I}. In figure 2(b), we can see that the $\Delta = |\phi_\Omega - \rho_{\mathrm{I}}|$ and $\overline{\Delta} = \left|\overline{\phi_\Omega} - \rho_{\mathrm{I}}\right|$ are so minute that they have little effect on infection risk analysis only. Notably, the results of ϕ_Ω are almost consistent with our simulation results, which verifies the correctness of equation (5). Overall, we can conclude that $\overline{\phi_\Omega}$ provides an accurate estimation of GHI at acceptable running times when $\Omega \ll N$.

In general, the number of initial infection sources Ω is unknown, but the initial infection proportion of the epidemic in the population, denoted by I_0, can generally be estimated from statistical sampling or clinical data. When I_0 is known, equation (10) can be extended to equation (11). Different from the precise number of infection sources Ω in a single outbreak, $N \times I_0$ represents the expectation of the number of

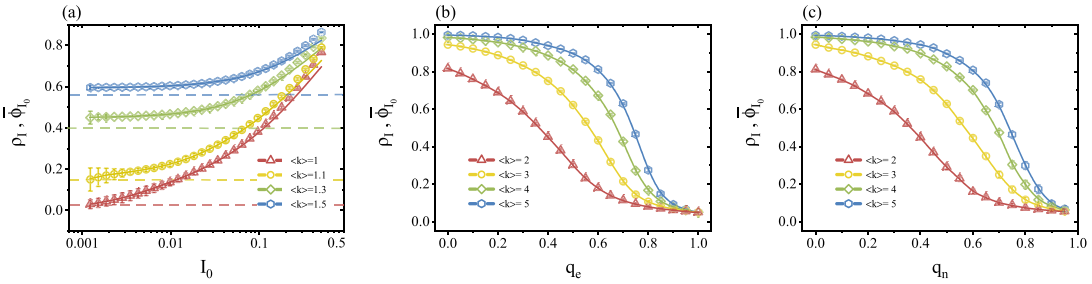

Figure 3. The performance of $\overline{\phi_{I_0}}$ on ER networks with $N = 10\,000$ and different average degree $\langle k\rangle$. The ER networks are the same for a fixed $\langle k\rangle$ in different experiments. (a) Comparison of $\overline{\phi_{I_0}}$ (lines) and ρ_I (symbols) with a different I_0, where ρ_I is the average fraction of infected nodes obtained from simulations of the SI model. When I_0 is less than about 0.2, the $\overline{\phi_{I_0}}$ tends to have the same value as the spreading simulation. The dashed lines in this graph are P_∞ for each ER network, where P_∞ is the proportion of nodes in the GCC. Approximate estimations of the ρ_I using the $\overline{\phi_{I_0}}$ under $I_0 = 5\%$ in the ER network with randomly removed edges (b) and nodes (c), where q_e is the proportion of edges removed and q_n is the proportion of nodes removed.

infection sources when each node is initially infected with the probability I_0.

$$\overline{\phi_{I_0}} = \sum_{1\leqslant i\leqslant L} p_i \left(1 - \left(1 - p_i\right)^{N\times I_0}\right). \tag{11}$$

To further verify that $\overline{\phi_{I_0}}$ can effectively estimate GHI, $\overline{\phi_{I_0}}$ is compared with the infection's prevalence ρ_I in an epidemic under the SI model for the networks with different component distributions. In figure 3(a), we generate disconnected ER networks with different numbers and sizes of connected components by adjusting the average degree $\langle k\rangle$. Moreover, we randomly remove the edge (in figure 3(b)) or nodes (in figure 3(c)) in the ER network to create different component distributions. The results in figure 3 shows that the simulations (symbols) are in excellent agreement with the theoretical results (lines). These results indicate that $\overline{\phi_{I_0}}$ can reasonably estimate the GHI, which is especially true when $I_0 \leqslant 0.1$. With a low level of time complexity and effective estimates of GHI, an efficient immunization strategy aims to reduce the indicator $\overline{\phi_{I_0}}$ of the network after immunization.

4. Efficient immunization strategy based on the GHI

4.1. The optimization model of the immunization strategy

Vaccination is one of the most effective ways to prevent or suppress the spread of an epidemic. From the viewpoint of vaccination, immunization corresponds to an attack that destroys the network on which it could spread. This paper considers node immunization approaches and assumes that the attached spreading edges are removed if a node is immunized. The set of immunized nodes is denoted by V_{immu}. The number of immunized nodes is denoted by n, and $p = n/N$ is the immunized proportion of the nodes. An immunization strategy is defined by $\boldsymbol{X} = (x_1, x_2, \ldots, x_N)$, where $x_j = 0$ if $v_j \in V_{\text{immu}}$, otherwise $x_j = 1$. Thus, we obtain the number of immunized nodes $n = N - \sum_{j=1}^{N} x_j$. The goal of our immunization method is to identify the optimal solution $\boldsymbol{X}^*$ which could minimize the GHI of the network after immunization. With the knowledge of I_0, we define $\Phi_{\text{GHI}}(\boldsymbol{X})$ as the $\overline{\phi_{I_0}}$ of the network immunized by strategy $\boldsymbol{X}$. We introduce an optimization model to solve the immunization strategy, which can be described as

$$\min\ \Phi_{\text{GHI}}\left(\boldsymbol{X} = (x_1, x_2, \ldots, x_N)\right)$$
$$\text{s.t.} \begin{cases} n = N - \sum_{j=1}^{N} x_j \\ x_j = 0 \text{ or } 1 \\ j = 1, 2, \ldots, N \end{cases} \tag{12}$$

where $j = 1, \ldots, N$ represents the serial number of nodes, and $\Phi_{\text{GHI}}(\boldsymbol{X})$ is used as the objective function of the optimization model to measure the effect of $\boldsymbol{X}$. The solution of the optimization model determines the optimal GHI strategy.

As a contrast, the optimal GCC strategy replaces the objective function $\Phi_{\mathrm{GHI}}(\boldsymbol{X})$ in equation (12) with $\Phi_{\mathrm{GCC}}(\boldsymbol{X})$ to minimize the GCC of the network, where $\Phi_{\mathrm{GCC}}(\boldsymbol{X})$ is the size of the GCC in the network after immunization. Meanwhile, we also compare the optimal GHI strategy with mainstream strategies, including a high-degree adaptive (HDA) strategy and the CI strategy [31]. The HDA strategy removes the nodes according to the adaptive computation of the degree. The CI strategy iteratively calculates the CI value of nodes and removes the node with the highest CI value. The CI value is an extension of the degree centrality which concerns the neighbors of node v_j at a distance of ℓ and that was set at $\ell = 2$ in this paper.

4.2. Experimental design

4.2.1. Tabu search algorithm

The tabu search algorithm [36, 45] has been proved to be an effective method for solving similar problems in the network and thus has been applied here to seek the optimal solution for the above optimization model. The basic principle of the tabu search is to pursue an optimal solution whenever it encounters a local optimum by allowing non-improving moves. Cycling back to previously visited solutions is prevented by using memories, called tabu lists, that record the search's recent history. The procedure of the algorithm is described below.

Step 1: initialization. We set the length of the tabu list $L_{\mathrm{tabu}} = 100$, the number of candidates $n_{\mathrm{can}} = 500$, the maximum total iteration number $T_{\max} = 30\,000$, the maximum iteration number without improvement of solution $n_{\max} = 5000$. The termination condition of the algorithm is when the present iteration step T_{iter} reaches $T_{\max}$ or the number of iterations for which the optimal solution is not updated n_{iter} exceeds $n_{\max}$.

Step 2: generate the initial solution $\boldsymbol{X}_0$. $\boldsymbol{X}_0$ can either be given randomly or by another strategy with a better performance. Let the current best solution $\boldsymbol{X}_{\mathrm{opt}} = \boldsymbol{X}_0$. Calculate $\Phi(\boldsymbol{X}_{\mathrm{opt}})$.

Step 3: determine the termination condition. If $T_{\mathrm{iter}} > T_{\max}$ or $n_{\mathrm{iter}} > n_{\max}$, the process stops and output $\boldsymbol{X}_{\mathrm{opt}}$ as the results; otherwise, continue to step 4.

Step 4: generate candidate solution. Generate n_{can} new candidate solutions $\boldsymbol{X}_{\mathrm{can}}$ by swapping the state of two nodes randomly. Determine $\boldsymbol{X}_{\mathrm{now}}$ by $\boldsymbol{X}_{\mathrm{now}} = \max \Phi(\boldsymbol{X}_{\mathrm{can}})$.

Step 5: update the tabu list. Determine whether $\boldsymbol{X}_{\mathrm{cur}} \notin T_{\mathrm{list}}$ or $\Phi(\boldsymbol{X}_{\mathrm{cur}}) < \Phi(\boldsymbol{X}_{\mathrm{opt}})$ (aspiration criterion). If satisfied, add $\boldsymbol{X}_{\mathrm{cur}}$ to T_{list}. If not satisfied, find another $\boldsymbol{X}_{\mathrm{cur}}$ s.t. $\boldsymbol{X}_{\mathrm{cur}} = \max \Phi(\boldsymbol{X}_{\mathrm{opt}})$ and $\boldsymbol{X}_{\mathrm{cur}} \notin T_{\mathrm{list}}$, and then add $\boldsymbol{X}_{\mathrm{cur}}$ to T_{list}. Notably, all the elements in the tabu list are abandoned in a certain number of iterations L_{tabu}.

Step 6: update the current best solution $\boldsymbol{X}_{\mathrm{opt}}$. Determine whether $\Phi(\boldsymbol{X}_{\mathrm{cur}}) < \Phi(\boldsymbol{X}_{\mathrm{opt}})$. If satisfied, then $\Phi(\boldsymbol{X}_{\mathrm{opt}}) = \Phi(\boldsymbol{X}_{\mathrm{cur}})$, $T_{\mathrm{list}} = \mathrm{NULL}$. If not satisfied, then return to step 3.

After obtaining the approximate optimal solution, a set of nodes is identified whose removal from the network can minimize $\Phi(\boldsymbol{X})$. The optimal GHI and the optimal GCC strategies are obtained by using the objective function $\Phi_{\mathrm{GHI}}(\boldsymbol{X})$ and $\Phi_{\mathrm{GCC}}(\boldsymbol{X})$, respectively.

4.2.2. Networks

Many social networks conform to the typical characteristics of small-world, scale-free (SF), or community structures. Hence, we analyze the case of three basic model networks, the Watts–Strogatz (WS) network [46], the SF network [47], and the KOSKK network [48, 49].

The WS model starts from a ring of $N = 1000$ vertices, each of which symmetrically connects to its four nearest neighbors. Then, a fraction of the edges in the network are rewired by visiting all four clockwise edges of each vertex and reconnecting them, with probability $p_{\mathrm{re}} = 0.5$, to a randomly chosen node.

The SF network is generated using preferential attachment [50], which signifies that the more connected a node is, the more likely it is to receive new links. The preferential attachment model is initiated with a small nucleus of $m_0 = 5$ fully connected nodes. Then, at every time step, a new node is added, with $m = 4$ links connected to an old node v_j whose degree is k_j with the probability equal to $k_j / \sum_j k_j$.

The KOSKK model is a dynamic network evolution model [48, 49] that can generate networks with typical features of social networks by utilizing network link weights. The network is initiated with N nodes and zero edges, and then evolved with three mechanisms:

(a) Local attachment. Select a node v_j randomly, and choose one of neighbor v_k with probability $\omega_{jk} / \sum_k \omega_{jk}$, where ω_{jk} is the weight on link e_{jk}. If v_j has another neighbor, choose one of them with probability $\omega_{kl} / \sum_l (\omega_{kl} - \omega_{jk})$. If there is no link between v_j and v_l, connect v_j and v_l with probability p_Δ and set the weight of new link as w_0. Increase link weight by δ.

(b) Global attachment. Connect v_j to a random node with probability p_r (or with probability 1 if v_j has no connections) and set the weight of new link as w_0.

(c) Node deletion. Select a random node and with probability p_d remove all of its connections.

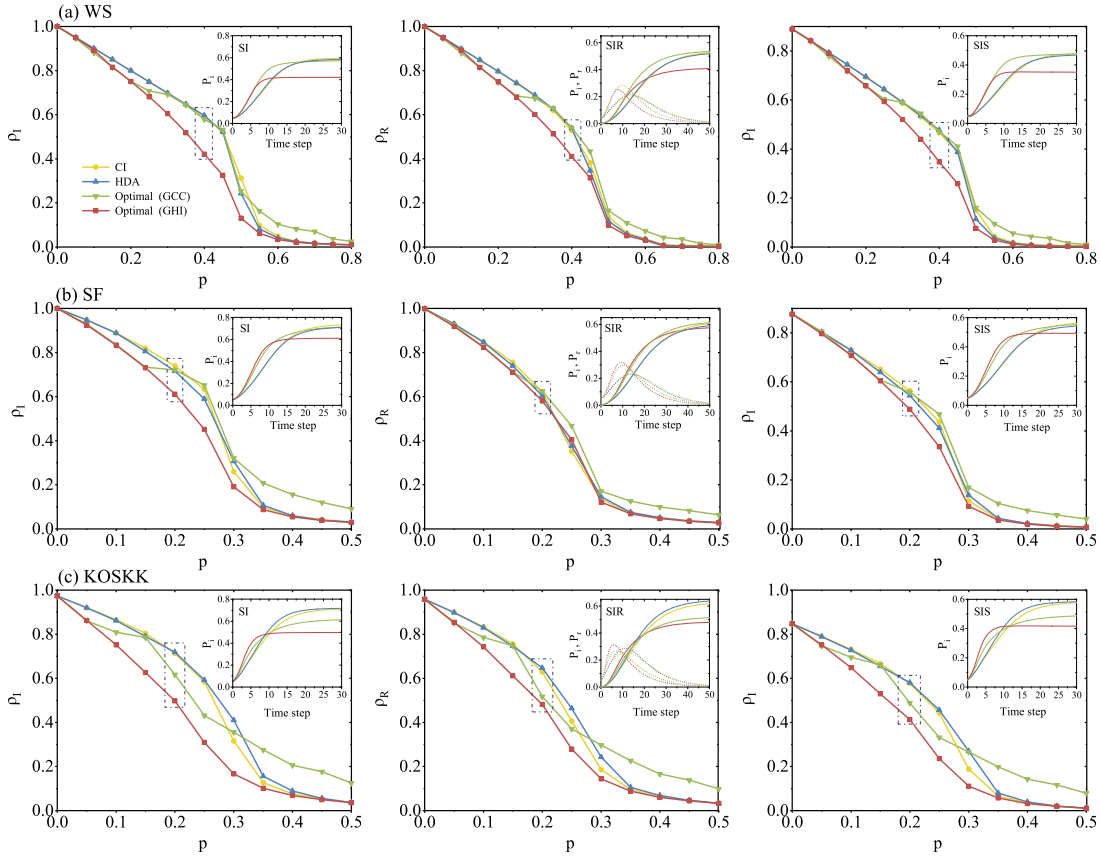

Figure 4. Comparisons of the immunization strategies HDA (blue line), CI (yellow line), optimal GCC (green line), and optimal GHI (red line) for several artificial networks: (a) the WS network, (b) the SF network, and (c) the KOSKK network. Plotted is the fraction of infected ρ_{I} or recovered individuals ρ_{R} at the steady state as a function of the immunized node fraction p in the SI, SIR, and SIS models. In the insets, temporal evolution of the spread process for the point in the dotted–dashed box are shown, where P_{i} is the fraction of infected nodes and P_{r} is the fraction of recovered nodes. The experiments set the infection rate $\beta = 0.25$, the recovery rate $\mu = 0.1$, and $I_0 = 5\%$.

The network is obtained after 10^7 time steps evolution, and the parameters are set as $N = 1000$, $\omega_0 = 1$, $p_r = 0.005$, $p_d = 0.001$, $p_\Delta = 0.25$, and $\delta = 0.6$.

4.2.3. Simulations of epidemic spread

The SI model is somewhat of an oversimplification that is valid only in cases where the time scale of recovery is much longer than the time scale of infection. More realistic models have been proposed in order to better accommodate the biological properties of real diseases. For instance, the susceptible–infectious–susceptible (SIS) and the susceptible–infectious–recovery (SIR) epidemiological models [1, 43]. To study the optimal GHI immunization strategy, we compared its efficiency with other strategies in the SIS and SIR models. The comparison results are given in figures 4 and 5. The SIS and SIR models are widely used to simulate the spread of epidemics in a network. In the SIS and SIR models, each node of the network represents an individual, and each edge is a connection through which the infection can spread. In the simulations of this paper, the SIS and SIR spreading processes are implemented by using synchronous updating methods. Namely, at each time step, each susceptible node is infected by its infected neighbor (the node connected) with probability β (infection rate) if it is connected to one or more infected nodes. At the same time, all infected nodes recover with probability μ (recovery rate). The dynamical process terminates when the system reaches a steady state. The SIR model assumes that an infectious individual who recovers from the disease has acquired permanent immunity. In the SIR model, the infection will eventually die out. Conversely, the SIS model assumes that the disease does not confer immunity so that individuals can be infected over and over again. Under SIS, the disease can reach a steady state, where a certain fraction of the population are kept infected. Considering this difference, when we measure the result of the SIR model, the fraction of individuals who have ever caught the disease is denoted by ρ_{R}. For the SIS model, it is the fraction of infected nodes persisting in the steady state denoted by ρ_{I}. p is defined as the fraction of immunized nodes.In the simulation, each node is initially infected with the probability $I_0 = 5\%$ (independent of the other nodes), and the spreading model starts with the parameters $\beta = 0.25$ and $\mu = 0.1$, averaging over 10 000 independent runs.

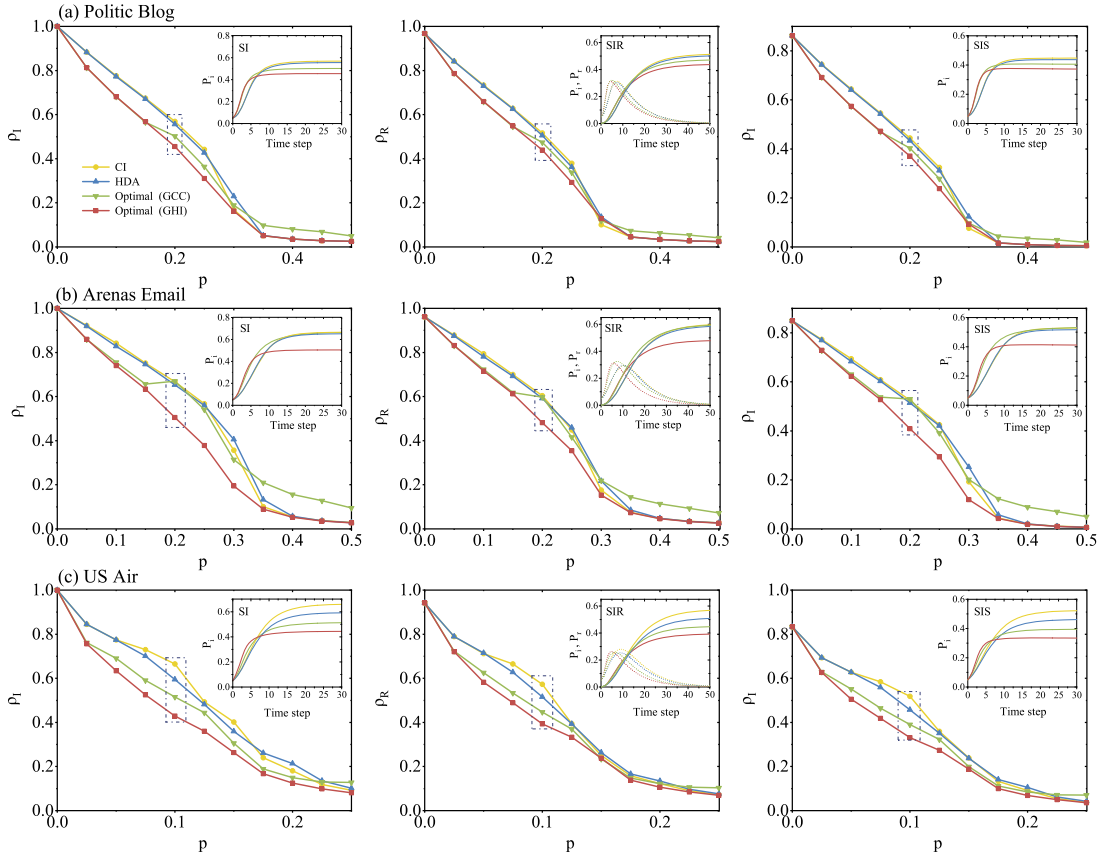

Figure 5. Comparison of the optimal GHI strategy and other immunization strategies for several real-world networks: (a) the politic blog network, (b) the Arenas email network, and (c) the US air transportation network. The experiments set the infection rate $\beta = 0.25$, the recovery rate $\mu = 0.1$, and $I_0 = 5\%$.

Table 1. The basic topological features of the networks. N and W are the number of nodes and links, where $\beta_{th} = \langle k \rangle / \langle k^2 \rangle$ is the epidemic threshold of a network, and $\langle k \rangle$ and $\langle k^2 \rangle$ are the mean degree and second-order mean degree of a network, respectively. C is the average clustering coefficient of a network.

Network	N	W	$\langle k \rangle$	β_{th}	C
WS	1000	4000	8	0.1196	0.0907
SF	1000	3904	7.81	0.0508	0.0436
KOSKK	1000	4474	8.95	0.0699	0.5532
Politic blog	1222	16 714	27.36	0.0123	0.3203
Arenas email	1133	5451	9.62	0.0535	0.2202
US air	332	2126	12.81	0.0225	0.6252

4.3. Results in synthetic and real networks

To study the efficiency of the optimal GHI strategy, we focused on the fraction of infected nodes ρ_I (steady state) in the SI and SIS models, and the fraction of recovery nodes ρ_R (steady state) in the SIR model. Smaller ρ_I or ρ_R indicates higher efficiency of a strategy. In the simulations, we look at the ρ_I or ρ_R in the stationary regime (steady state) as a function of the fraction of immunized nodes p. We implement the optimal GHI strategy and other strategies to immunize p proportions of individuals in the networks. Then we let each node is initially infected with the probability $I_0 = 5\%$, and iterate the SI, SIS, and SIR infection process with synchronous updating. The SI, SIS, and SIR process are implemented with a fixed infection rate $\beta = 0.25$, and the SIS, SIR process fix the recovery rate $\mu = 0.1$. After the system reaches the steady state, ρ_I or ρ_R is obtained.

The results of the model networks shown in figure 4 reveal that the optimal GHI strategy has better performance than other strategies, especially in the WS and KOSKK networks. Meanwhile, the advantage of the optimal GHI strategy is not evident in the SF network for the SIR model, and the effects of the HDA and CI strategies are close to that of the optimal GHI strategy. The temporal evolution of spread process in

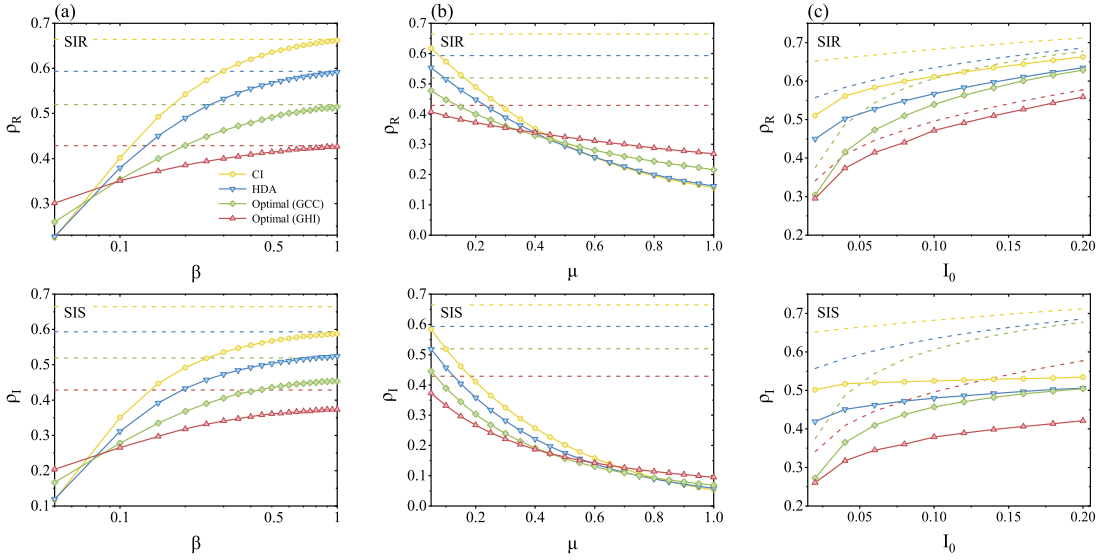

Figure 6. Comparisons of the optimal GHI strategy and other immunization strategies under different parameters in the SIR and SIS models. Plotted is the fraction of infected individuals ρ_{I} (steady state) and the fraction of recovered individuals ρ_{R} (steady state) versus the infection rate β (a), the recovery rate μ (b), and initial infection proportion I_0 (c). The experiments immunized a nodes fraction $p = 0.1$ of the US air transportation network and set the control variables as $\beta = 0.25$, $\mu = 0.1$, and $I_0 = 5\%$. The dashed lines in this graph are the $\overline{\phi_\Omega}$ of networks after immunization for each strategy.

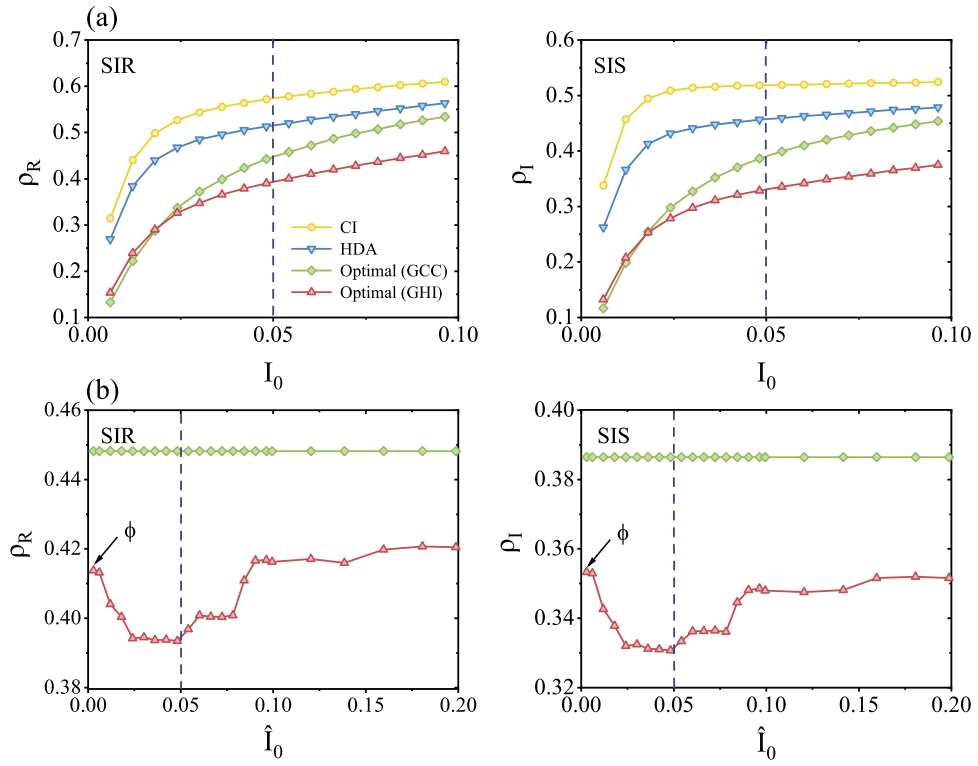

Figure 7. Performance of the optimal GHI strategy with a certain missing accuracy of I_0 on the US air transportation network: (a) comparison of the optimal GHI strategy obtained by $\hat{I}_0 = 5\%$ with other strategies for simulations with different I_0. (b) The simulation results for different optimal GHI strategies obtained through different $\hat{I}_0$ under the same $I_0 = 5\%$. The dashed lines represent the case of $\hat{I}_0 = I_0$.

networks after immunization are given in the insets. For all networks, the infected fraction is significantly lower when using the optimal GHI strategy as compared to other strategies with the same fraction of immunization doses.

The model networks cannot fully describe the characteristics of the real systems. Therefore, we also implemented the selected strategies for three real-world network examples through which epidemics are

spread: the politic blog network [51], the Arenas email network [52], and the US air transportation network [53]. The details of these networks are given in table 1. Some conclusions obtained from the model networks are also shown in the real networks. The experiments demonstrate that the optimal GHI strategy exhibits a clear advantage with fewer nodes immunized to achieve the same immunization effect when compared to other targeted strategies (figure 5). In addition, the fraction of infected individuals for the optimal GHI strategy is significantly lower than those for other strategies with the same fraction of immunization doses. These results show that the optimal GHI strategy for SF characteristics in real networks, and reducing the size of the GCC (P_∞) is not as effective as reducing the GHI in the network.

The network is fragmented into many connected components of different sizes by immunization. The size distribution of connected components plays a more significant role than the topology in components for these networks. The immunization strategies with different mechanisms make the distribution of connected components of the network after immunization different. GHI is used to evaluate the infection risk of the network after immunization based on the distribution of connected components. Therefore, the optimal GHI strategy, which minimizes the GHI by immunizing nodes, shows great advantages on immunization in different networks through other strategies.

4.4. Robustness of the optimal GHI strategy

So far, we have focused on the performance of the optimal GHI strategy in different networks. These results suggest that, although GHI cannot accurately quantify the expected infection's prevalence under the SIS and SIR models, GHI reflects the structural connectivity of the network and quantizes the impact of the distribution of connected components to the spreading. However, the parameters of the dynamic models are also factors which affect the result of simulation. In previous experiments, the parameters of the SIS and SIR models are fixed. Next, we need to verify whether the strategy is effective under different parameters in the epidemiological model. In this section, we move our focus to the robustness of the optimal GHI strategy and define the robustness in two ways. On the one hand, we consider the robustness of the strategy in terms of sensitivity to infectious disease model parameters. On the other hand, we also evaluate the robustness against the deviation of prior information in the sense of how well the optimal GHI strategy yields, even when the estimated value we obtained does not strictly agree with the precise I_0.

To further clarify what types of infectious diseases the optimal GHI strategy is suitable for, the performance of different strategies is compared under different infectious disease model parameters on the US air transportation network. There, an infected airport implies that sick people arrive or depart from it. Consequently, immunization means all people at an airport are screened, flights are canceled, or the entire airport is shut down. In figure 6(a), we immunized a fraction $p = 0.1$ of nodes and compared the impact of different infection rates β on the efficiency of the optimal GHI strategy and at a fixed recovery rate $\mu = 0.1$. Similar experiments were done under different recovery rates μ, at a fixed infection rate $\beta = 0.25$ (figure 6(b)). When β is high or μ is low, the proportion of infections under the optimal GHI strategy maintains a relatively low level. As β increases or μ decreases, the infection's prevalence in the simulation is tend to the value of GHI in the network, and the advantages of the optimal GHI strategy increase significantly. The results show that the effect of the optimal GHI strategy is pronounced for infectious diseases which are highly contagious and difficult to recover. Moreover, from the simulation result in figures 6(a) and (b), the optimal GHI strategy which is effective in the SI model is similarly effective for SIS and SIR models. The effective parameter range of the optimal GHI strategy is suitable for the infection and recovery rates of many real infectious diseases, e.g., SARS and COVID-19.

In addition, we tested the effects of the optimal GHI strategy to deal with the different initial infection proportions I_0 shown in figure 6(c). It can be seen that the reduced prevalence when the optimal GHI strategy is used performs much better than other strategies under different I_0.

Meanwhile, in a real-world situation, there is typically no access to the precise initial infection proportion (denoted by I_0), and estimated values of I_0 (denoted by $\hat{I}_0$) are generally used to guide decisions. Meanwhile, the initial sources of infection are also not randomly generated and present correlation and aggregation. These factors imply that the $\hat{I}_0$ referred to for making decisions has a certain deviation from I_0. Thus far, we have conducted simulations with the assumption that we know a precise I_0, which is the ideal case for $\hat{I}_0 = I_0$. Now, we consider how robust the optimal GHI strategy is against the noise of I_0. In the experiment, $\hat{I}_0$ is the initial infection proportion used to formulate the optimal GHI strategy, and I_0 is the initial infection proportion used in the simulations of the SIS and SIR model. Without the knowledge of the real I_0, we formulate an optimal GHI strategy based on the estimated $\hat{I}_0$ value. To determine whether this strategy is still valid in the simulation with real I_0, we test the effectiveness of the strategy in simulation experiments with different I_0. As is shown in figure 7(a), we take $\hat{I}_0 = 5\%$ to obtain the optimal GHI strategy and study the effects of the optimal GHI strategy for simulations under different I_0. We find that the epidemic can still be more effectively controlled by the optimal GHI strategy than by others given the

same conditions. Moreover, to further test the impact of the estimation accuracy of I_0 on the effect of the optimal GHI strategy, we fixed the I_0 used in the simulation to test the effect of strategies obtained by different estimated values $\hat{I}_0$. Figure 7(b) shows the effects of the optimal GHI strategy obtained with different $\hat{I}_0$ for the simulation under $I_0 = 5\%$. It suggests that our optimal GHI strategy still maintains an effective performance even if there is a certain deviation between the $\hat{I}_0$ we obtained and the actual value I_0. As the estimated $\hat{I}_0$ moves closer to the actual I_0, the effect of the optimal GHI strategy becomes increasingly evident.

Based on these experiments, we conclude that the optimal GHI strategy exhibits a low sensitivity to the epidemiological model parameters and a certain robustness against the noise of I_0.

5. Conclusion and discussion

In this paper, we proposed the indicator named GHI to measure the infection risk of individuals in a network according to the number of infection sources, along with a computationally efficient method to approximate it. We set our immunization goal as minimizing GHI and established an optimization model to search for the immunization strategy. Our method can immunize or quarantine the population against possible multi-regional outbreaks based on initially infected proportions. We discussed extensive experiments on both synthetic and real-world networks using SIS and SIR simulations. The results show that the optimal GHI method is significantly more efficient at preventing the spread of disease spreading than other basic immunization methods, especially with highly infectious or low recovery rate diseases. Moreover, our strategy shows a certain robustness for deviations in this prior information, which makes the method suitable for the requirements of practical applications.

GHI measures the upper bound of the expected fraction of infected nodes, which is based on the strong assumption that all nodes in the connected components containing infected nodes are at risk of infection. There are many different methods to minimize GHI besides immunization of nodes, such as immunization of links and community isolation. Further research may consider the following two aspects. On the one hand, GHI can be introduced to more application scenarios, and its properties need to be explored further. On the other hand, the tabu search algorithm we utilized in this study is computationally expensive and complex in the face of certain large-scale networks. Therefore, it is necessary to explore heuristics to reduce the computational complexity.

Acknowledgments

Xiaojun Duan is supported by the National Natural Science Foundation of China under Grant No. 11771450. Xin Lu is supported by the National Natural Science Foundation of China (91846301, 71790615, 72025405, 82041020 and 71771213) and the Hunan Science and Technology Plan Project (2019GK2131, 2020TP1013). Jürgen Kurths is supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Center 'Digital biodesign and personalized healthcare' No.075-15-2020-926. Mingze Qi is supported by the Postgraduate Scientific Research Innovation Project of Hunan Province (CX20200001). This study is also partially supported by the Shenzhen Basic Research Project for Development of Science and Technology (JCYJ20200109141218676).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Peng Chen https://orcid.org/0000-0001-8240-8322
Mingze Qi https://orcid.org/0000-0001-9552-6504
Xin Lu https://orcid.org/0000-0002-3547-6493

References

[1] Anderson R M and May R M 1991 *Infectious Diseases of Humans: Dynamics and Control* (Oxford: Oxford University Press)
[2] Diekmann O and Heesterbeek J A P 2000 *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (New York: Wiley)

[3] Hethcote H W 2000 *SIAM Rev.* **42** 599–653
[4] Watts D J, Peretti J and Frumin M 2007 *Viral Marketing for the Real World* (Boston: Harvard Business School Publication)
[5] Wolfram C 2020 *Complex Syst.* **29** 87–105
[6] Gallotti R, Valle F, Castaldo N, Sacco P and Domenico M D 2020 *Nature Human Behaviour* **4** 1285–93
[7] Borowiak M, Ning F, Pei J, Zhao S, Tung H R and Durrett R 2020 (arXiv:2008.07293)
[8] Chang S, Pierson E, Koh P W, Gerardin J, Redbird B, Grusky D and Leskovec J 2021 *Nature* **589** 82–7
[9] Newman M E J, Barabási A L and Watts D J 2006 *The Structure and Dynamics of Networks* (Princeton, NJ: Princeton University Press)
[10] Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A 2015 *Rev. Mod. Phys.* **87** 925–79
[11] Wang Z, Bauch C T, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, Perra N, Salathé M and Zhao D 2016 *Phys. Rep.* **664** 1–113
[12] Chen X, Wang R, Tang M, Cai S, Stanley H E and Braunstein L A 2018 *New J. Phys.* **20** 13007
[13] Muro M A D, Alvarez-Zuzek L G, Havlin S and Braunstein L A 2018 *New J. Phys.* **20** 83025
[14] Newman M E J 2005 *Soc. Network.* **27** 39–54
[15] Masuda N 2009 *New J. Phys.* **11** 123018
[16] Zeng A and Zhang C-J 2013 *Phys. Lett.* A **377** 1031–5
[17] Wang Z, Zhao Y, Xi J and Du C 2016 *Phys.* A **461** 171–81
[18] Cohen R, Havlin S and ben-Avraham D 2003 *Phys. Rev. Lett.* **91** 247901
[19] Holme P 2004 *Europhys. Lett.* **68** 908–14
[20] Chen S and Lu X 2017 *Sci. Rep.* **7** 3268
[21] Altarelli F, Braunstein A, Dall'Asta L, Wakeling J R and Zecchina R 2014 *Phys. Rev.* X **4** 21024
[22] Li S, Zhao D, Wu X, Tian Z, Li A and Wang Z 2020 *Appl. Math. Comput.* **366** 124728
[23] Chen Y, Paul G, Havlin S, Liljeros F and Stanley H E 2008 *Phys. Rev. Lett.* **101** 58701
[24] Clusella P, Grassberger P, Pérez-Reche F J and Politi A 2016 *Phys. Rev. Lett.* **117** 208301
[25] Liu Y, Wang X and Kurths J 2019 *IEEE Trans. Evol. Comput.* **23** 1049–63
[26] Liu Jian-Guo J, Ren Zhuo-Ming Z, Guo Qiang Q and Wang Bing-Hong B 2013 *Acta Phys. Sin.* **62** 178901
[27] Wang J, Li C and Xia C 2018 *Appl. Math. Comput.* **334** 388–400
[28] Li C, Wang L, Sun S and Xia C 2018 *Appl. Math. Comput.* **320** 512–23
[29] Lü L, Chen D, Ren X-L, Zhang Q-M, Zhang Y-C and Zhou T 2016 *Phys. Rep.* **650** 1–63
[30] Holme P, Kim B J, Yoon C N and Han S K 2002 *Phys. Rev.* E **65** 56109
[31] Morone F and Makse H A 2015 *Nature* **527** 544
[32] Braunstein A, Dall'Asta L, Semerjian G and Zdeborová L 2016 *Proc. Natl Acad. Sci. USA* **113** 12368–73
[33] Zdeborová L, Zhang P and Zhou H-J 2016 *Sci. Rep.* **6** 37954
[34] Ren X-L, Gleinig N, Helbing D and Antulov-Fantulin N 2019 *Proc. Natl Acad. Sci. USA* **116** 6554–9
[35] Tian L, Bashan A, Shi D-N and Liu Y-Y 2017 *Nat. Commun.* **8** 14223
[36] Deng Y, Wu J and Tan Y-j. 2016 *Phys.* A **442** 74–81
[37] Qi M, Deng Y, Deng H and Wu J 2018 *Chaos* **28** 121104
[38] Fan C, Zeng L, Sun Y and Liu Y-Y 2020 *Nature Machine Intelligence* **2** 317–24
[39] Nishi A *et al* 2020 *Proc. Natl Acad. Sci. USA* **117** 30285–94
[40] Costa G S and Ferreira S C 2020 *Phys. Rev.* E **101** 22311
[41] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D 2006 *Phys. Rep.* **424** 175–308
[42] Newman M 2010 *Networks: An Introduction* (Oxford: Oxford University Press)
[43] Daley D J and Gani J M 1999 *Epidemic Modelling: An Introduction* (Cambridge: Cambridge University Press)
[44] Rhoades S A 1993 *Fed. Reserv. Bull.* **91** 188–9
[45] Glover F and Laguna M 1997 *Tabu Search* (NewYork: Springer)
[46] Watts D J and Strogatz S H 1998 *Nature* **393** 440–2
[47] Barabási A-L and Albert R 1999 *Science* **286** 509–12
[48] Kumpula J M, Onnela J-P, Saramäki J, Kaski K and Kertész J 2007 *Phys. Rev. Lett.* **99** 228701
[49] Lu X 2013 *Soc. Network.* **35** 669–85
[50] Albert R Z and Barabási A L 2001 *Rev. Mod. Phys.* **74** 47–97
[51] Adibi J, Grobelnik M, Mladenic D and Pantel P 2005 *Kdd'05 the Eleventh Acm Sigkdd Int. Conf. on Knowledge Discovery and Data Mining*
[52] Guimerà R, Danon L, Díaz-Guilera A, Giralt F and Arenas A 2003 *Phys. Rev.* E **68** 65103
[53] Batagelj V and Mrvar A 1998 *Pajek-program for large Network analysis Connections* **21** 47–57