

## Research Article

# Group-Based Susceptible-Infectious-Susceptible Model in Large-Scale Directed Networks

Xu Wang<sup>1,2</sup>, Bo Song<sup>2,3</sup>, Wei Ni<sup>4</sup>, Ren Ping Liu<sup>2</sup>, Y. Jay Guo<sup>2</sup>,  
Xinxin Niu<sup>1,5</sup> and Kangfeng Zheng<sup>1</sup>

<sup>1</sup>School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>2</sup>Global Big Data Technologies Centre, University of Technology Sydney, 2007, Australia

<sup>3</sup>School of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, 210003, China

<sup>4</sup>Data 61, CSIRO, Sydney, NSW 2122, Australia

<sup>5</sup>State Key Laboratory of Public Big Data, Guizhou 550025, China

Correspondence should be addressed to Xu Wang; [xu.wang-3@student.uts.edu.au](mailto:xu.wang-3@student.uts.edu.au)

Received 17 October 2018; Accepted 8 January 2019; Published 16 January 2019

Academic Editor: Angel M. Del Rey

Copyright © 2019 Xu Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Epidemic models trade the modeling accuracy for complexity reduction. This paper proposes to group vertices in directed graphs based on connectivity and carries out epidemic spread analysis on the group basis, thereby substantially reducing the modeling complexity while preserving the modeling accuracy. A group-based continuous-time Markov SIS model is developed. The adjacency matrix of the network is also collapsed according to the grouping, to evaluate the Jacobian matrix of the group-based continuous-time Markov model. By adopting the mean-field approximation on the groups of nodes and links, the model complexity is significantly reduced as compared with previous topological epidemic models. An epidemic threshold is deduced based on the spectral radius of the collapsed adjacency matrix. The epidemic threshold is proved to be dependent on network structure and interdependent of the network scale. Simulation results validate the analytical epidemic threshold and confirm the asymptotical accuracy of the proposed epidemic model.

## 1. Introduction

Epidemic models have been widely used to analyze sophisticated interactions in networks, e.g., virus attack and propagation in computer networks [1, 2], rumor spreading in social networks [3], and cascading failures [4]. As a basic epidemic model, the susceptible-infected-susceptible (SIS) model defines two node states (susceptible and infected) and captures two state transition processes, namely, the *infection* from infected nodes to susceptible nodes and the *self-healing* of infected nodes. Earlier epidemic models omit topologies by assuming a network topology of a complete graph [5]. Network topologies, playing a key role in cyber security [6], are later considered and proved to have a strong impact on the epidemic propagation process [7].

Epidemic models trade the modeling accuracy for complexity reduction. Markov-chain based epidemic models are

able to precisely analyze the epidemic propagation process but require  $2^N$  Markov states to capture the S/I states of  $N$  nodes and therefore can hardly be applied to large-scale networks [7, 8]. A number of topological epidemic models [8–12] decompose the  $2^N$ -state Markov process into  $N$  small Markov processes. This is achieved by using the expected infection probability of every node instead of the actual node state. A significant result from the models is that the epidemic threshold, under which the epidemic will eventually die out, is given by  $1/\lambda_1(\mathbf{A})$ , where  $\lambda_1(\mathbf{A})$  is the largest eigenvalue of the adjacency matrix of the network topology [9].

Network features, e.g., the degree distribution of scale-free networks [13], can be employed to simplify epidemic models by adopting the degree-based mean-field approach [14]. Specifically, nodes with the same degree are assumed to be infected with the same probabilities. An interesting result of the epidemic in scale-free networks is the absence

of an epidemic threshold [15]. The degree-based mean-field approach has extended the SIS process, e.g., the epidemic propagation with incubation and the epidemic with a recovery state [16, 17]. However, the degree-based epidemic model cannot capture epidemic propagations in specific networks.

This paper presents a group-based continuous-time Markov model to quantitatively analyze the SIS process in large-scale directed networks. We start with the network modeling, where nodes are categorized into groups according to their connectivity. A collapsed adjacency matrix is proposed to describe the network topology. Based on the node groups, a continuous-time Markov model is proposed to capture the SIS-type propagation, where the state of a group is estimated by taking the mean-field approximation. Focusing on the problem of epidemic threshold, the proposed nonlinear model is linearized via omitting high-order terms around the disease-free point. The epidemic threshold is derived by performing matrix analysis on the Jacobian matrix of the linearized model and then validated by simulations. The key contributions of this paper can be summarized as follows:

- (i) We propose a new modeling method, which groups nodes with the same connectivity in directed networks and models the epidemic propagation of the groups by using continuous-time Markov SIS models.
- (ii) By taking the mean-field approximation, the proposed SIS model is asymptotically accurate with the decrease of effective spreading rate and/or the increase of node groups.
- (iii) Linearization and stability analysis are carried out on the proposed SIS model to deduce the epidemic threshold, under which the epidemic eventually becomes extinct.
- (iv) The epidemic threshold is proved to be dependent on network structure and interdependent of the network scale.

Comprehensive simulations confirm the validity of the proposed mean-field epidemic model and the deduced epidemic threshold in large-scale networks.

The rest of this paper is organized as follows. In Section 2, related works are reviewed. In Section 3, the directed network model is presented, followed by the proposed mean-field SIS model in Section 4. In Section 5, numerical and simulation results are provided, followed by conclusions in Section 6.

## 2. Related Work

The SIS model was firstly developed for biological infectious diseases, which defines two states for a node, i.e., susceptible or infected [5]. A susceptible node can be infected with probability  $\beta$  by an infected neighbor. An infected node can be cured with probability  $\gamma$ . Recently, various states were introduced into the SIS model. Kephart et al. [18] and Vojnovic et al. [19] proposed a new “warning” state at every node, in addition to the “infected” and “susceptible” states. The probability of a node switching to the “warning” state depends on the population of nodes in the warning state

and is independent of the network topology. The SIS model was extended with “quarantined”, “vaccinated”, and “delay” states, to capture the time-delayed worm propagation in computer networks [20]. These models significantly simplified the dynamic process by omitting the impact of network topologies.

The network topologies can have a strong impact on epidemic propagations [11]. In [8], the continuous-time Markov process was adopted to model epidemic propagations in specific topologies, where every Markov state collects the states of all nodes in the network. The second largest eigenvalue of the Markov transition matrix determines the convergence rate to the absorbing virus-free state. With a discrete-time Markov model, the convergence rate was proved to be asymptotically bounded [7]. The Markov models can be decomposed into  $N$  small Markov processes in  $N$ -node networks by applying the mean-field theory to achieve tractability [8, 10–12]. The decomposed  $N$  Markov processes deduced a widely approved result that a virus dies out quickly if  $\beta/\gamma < 1/\lambda_1(\mathbf{A})$ , where  $\mathbf{A}$  is the adjacency matrix of the network, and  $\lambda_1(\mathbf{A})$  is the largest eigenvalue of  $\mathbf{A}$ . The network can also be dynamic, where nodes can transfer among communities [21]. The simulation results on two communities revealed that the node mobility can accelerate the malware propagation and improve the epidemic threshold.

Some network features can be employed to simplify epidemic models and improve tractability. Pastor-Satorras et al. [15] derived the probability of a node being infected as a function of the expected number of its infected neighbors in scale-free networks. The number of infected nodes at the equilibrium state was given by  $\rho \approx 2e^{-1/m\tau}$ , where  $m$  is a network generation parameter and  $\tau$  is the effective spreading rate. Zou et al. [22] simulated the propagation of Internet email worms in scale-free graphs and showed that the aforementioned result can be overestimated due to the implicit homogeneous mixing assumption. Meanwhile, Li et al. showed the analysis on scale-free graphs is inaccurate for specific topologies [23].

The statistic topology models are generally based on undirected networks. However, there is an intrinsic directionality in the propagation in specific types of dynamics, e.g., infectious disease spreading [24] and information transmission [25]. Directed networks, sets of vertices, and a collection of directed edges that connect pairs of ordered vertices are useful to represent specific transmissions with intrinsic directionality in the propagation [14]. Meyers et al. [26] employed the percolation theory to predict disease transmission through semidirected contact networks, where edges may be directed or undirected and found that the probability of an epidemic and the expected fraction of a population infected during an epidemic can be different in semidirected networks, in contrast to the routine assumption that these two quantities are equal. Li et al. [27] defined the directionality  $\xi$  as the percentage of unidirectional links and found that the lower bound of the epidemic threshold increases with a growing  $\xi$ , implying that the directionality hinders the propagation of epidemic processes. In [28], Khanafer et al. studied the stability of an SIS  $N$ -intertwined Markov model over arbitrary directed network topologies

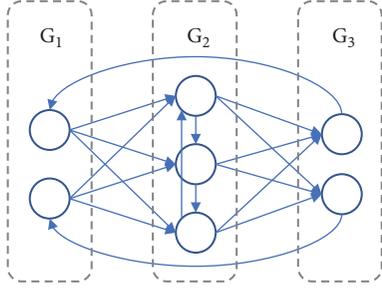


FIGURE 1: An example of networks we considered where 7 nodes are categorized into three groups according to the connection between them.

and showed that when the basic reproduction number is greater than one, the epidemic state is locally exponentially stable, and when the network is not initialized at the disease-free state, the epidemic state is globally asymptotically stable.

### 3. The Directed Network Model

We consider the SIS epidemic process in a strongly connected network with  $N$  nodes connected by directed edges. Each node in the network can be in either a susceptible ( $S$ ) or an infected ( $I$ ) state. An infected node can infect its susceptible neighbors along the directed edges at the rate of  $\beta$  per edge. Infected nodes can independently recover to be susceptible at the rate of  $\gamma$ .

We suppose that the  $N$  nodes in a network  $\mathcal{G}$  can be categorized into  $n$  groups, denoted by  $G_1, G_2, \dots, G_n$ , where the nodes in the same group have the same out-degrees and the same number of edges to the nodes in the same destination group. The number of  $G_i$ -nodes can be denoted by  $N_i$  (here,  $\sum_i N_i = N$ ). Given the node groups, we define a collapsed adjacency matrix, denoted by  $\mathcal{A}$ , to describe the topology of  $\mathcal{G}$ . The  $(i, j)$ -th entry of the matrix, denoted by  $a_{ij}$ , describes the number of edges from a  $G_i$ -node pointing to  $G_j$ -nodes. As a result,  $\mathcal{G}$  can be described by the node number vector  $\mathcal{N} = [N_1, N_2, \dots, N_n]$  and the collapsed adjacency matrix  $\mathcal{A}$ . Figure 1 provides an example of node categorization, where  $N = 7$  nodes are categorized into  $n = 3$  groups, i.e.,  $\mathcal{N} = [2, 3, 2]$ . Its collapsed adjacency matrix can be given by

$$\mathcal{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0 & 3 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \quad (1)$$

Other notations are defined as follows:  $[A]_i$  ( $A \in \{S, I\}$ ) denotes the number of  $G_i$ -nodes in state  $A$ .  $AB_{ij}$  denotes an edge starting from a  $G_i$ -node and ending at a  $G_j$ -node, where the  $G_i$ -node and the  $G_j$ -node are in states  $A$  and  $B$ , respectively.  $[AB]_{ij}$  denotes the number of  $AB_{ij}$  edges. Let  $[ABC]_{ijk}$  ( $[ABC]'_{ijk}$ ) denote the number of edge pairs consisting of  $AB_{ij}$  and  $BC_{jk}$  ( $AB_{ij}$  and  $CB_{kj}$ ), as illustrated

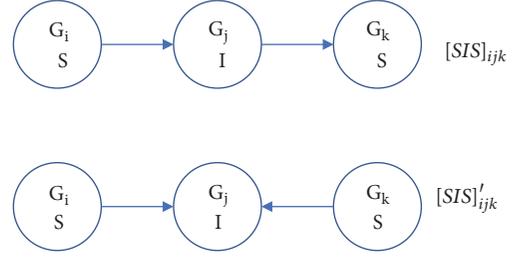


FIGURE 2: An example of  $[ABC]_{ijk}$  and  $[ABC]'_{ijk}$ .

by Figure 2. Numerical relationships between states of nodes and states of edges satisfy

$$N_i = [S]_i + [I]_i; \quad (2a)$$

$$a_{ij}N_i = [SS]_{ij} + [SI]_{ij} + [IS]_{ij} + [II]_{ij}; \quad (2b)$$

$$a_{ij}[S]_i = [SS]_{ij} + [SI]_{ij}; \quad (2c)$$

$$a_{ij}[I]_i = [IS]_{ij} + [II]_{ij}; \quad (2d)$$

$$[S]_i = N_i - [I]_i = N_i - \frac{1}{a_{ij}} ([IS]_{ij} + [II]_{ij}). \quad (2e)$$

Here, (2a) is because any node is in either an  $S$  or  $I$  state. (2b) is because any edge is in one of the four states given in the right-hand side (RHS) of (2b). Meanwhile, edges can be classified according to the state of the starting point, as given by (2c) and (2d). (2e) is deduced from (2a), (2b), and (2d).

### 4. Group-Based Mean-Field SIS Model

We propose to analyze the SIS process in directed networks by employing the mean-field approximation which uses a single average effect to approximate the effect of all the other individuals on any given individual. Thus, the same group of nodes in our model are evaluated by the same average estimation. The state transitions of nodes and edges in the SIS process can be given by

$$\frac{d[S]_i}{dt} = \gamma [I]_i - \sum_j \beta [IS]_{ji}; \quad (3a)$$

$$\frac{d[I]_i}{dt} = -\gamma [I]_i + \sum_j \beta [IS]_{ji}; \quad (3b)$$

$$\begin{aligned} \frac{d[SS]_{ij}}{dt} &= \gamma [SI]_{ij} + \gamma [IS]_{ij} - \sum_k \beta [ISS]_{kij} \\ &\quad - \sum_k \beta [SSI]'_{ijk}; \end{aligned} \quad (3c)$$

$$\begin{aligned} \frac{d[SI]_{ij}}{dt} &= -\gamma [SI]_{ij} + \gamma [II]_{ij} + \sum_k \beta [SSI]'_{ijk} \\ &\quad - \sum_k \beta [ISI]_{kij}; \end{aligned} \quad (3d)$$

$$\begin{aligned} \frac{d[IS]_{ij}}{dt} = & -\gamma [IS]_{ij} + \gamma [II]_{ij} - \beta [IS]_{ij} + \sum_k \beta [ISS]_{kij} \\ & - \sum_k \beta [ISI]_{ijk}'; \end{aligned} \quad (3e)$$

$$\begin{aligned} \frac{d[II]_{ij}}{dt} = & -2\gamma [II]_{ij} + \beta [IS]_{ij} + \sum_k \beta [ISI]_{kij} \\ & + \sum_k \beta [ISI]_{ijk}'. \end{aligned} \quad (3f)$$

Here, (3a) and (3b) give the changing rate of susceptible and infected  $G_i$ -nodes. The RHS is because an infected  $G_i$ -node can be cured with the rate  $\gamma$ ; a susceptible  $G_i$ -node can be infected with the rate  $\beta$  per edge by an infected  $G_j$  neighbor. In the continuous-time model, the time slot is infinitesimal that the infection rate can be summed together, i.e.,  $\sum_j \beta [IS]_{ji}$ .

Equations (3c)-(3f) capture the time-varying number of links. The first two terms on the RHS of (3c) are because an  $SS_{ij}$  edge can transfer from an  $SI_{ij}$  or  $IS_{ij}$  edge when the infected node is cured. The last two terms on the RHS of (3c) capture the cases where the starting  $G_i$ -node or the ending  $G_j$ -node is infected by its infected neighbors. (3d) is because an  $SI_{ij}$  edge can transfer to an  $SS_{ij}$  edge in the case that the infected  $G_j$ -node is cured at the rate  $\gamma$ ; and to  $II_{ij}$  in the case that the susceptible  $G_i$ -node is infected by its infected neighbors.  $SI_{ij}$  can transfer from  $II_{ij}$  in the case that the infected  $G_i$ -node is cured with the rate  $\gamma$  or from  $SS_{ij}$  in the case that the susceptible  $G_j$ -node is infected by its infected neighbors. Different from previous SIS models in undirected networks, e.g., [15],  $SI_{ij}$  cannot transfer to  $II_{ij}$ , as the epidemic can only propagate along directed edges. Equation (3e) can be similarly obtained attaching the infection process, i.e.,  $-\beta [IS]_{ij}$ . Equation (3f) is self-explanatory.

Known as the disease-free equilibrium point, the equilibrium point of interest is  $([SI]_{ij}, [IS]_{ij}, [II]_{ij}) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$ ; i.e., all nodes are susceptible. The condition of the disease-free equilibrium point can be deduced by linearizing the SIS model given by (3a)–(3f). This is because the stability of the original nonlinear system can be determined by the eigenvalues of the linearized model as stated in Lyapunov's First Method [29].

To linearize the SIS model, the number of edge pairs  $[ABC]_{ijk}$  and  $[ABC]_{ijk}'$  is first estimated by using the number of unpaired edges. This is achieved by applying the moment closure approximation as evaluated in [30, 31]. As a result, we have

$$[ISS]_{kij} = \frac{[IS]_{ki} [SS]_{ij}}{[S]_i} = \frac{[IS]_{ki} [SS]_{ij}}{N_i - ([IS]_{ip} + [II]_{ip})/a_{ip}}; \quad (4a)$$

$$[SSI]_{ijk}' = \frac{[IS]_{kj} [SS]_{ij}}{[S]_j} = \frac{[IS]_{kj} [SS]_{ij}}{N_j - ([IS]_{jq} + [II]_{jq})/a_{jq}}; \quad (4b)$$

$$[ISI]_{kij} = \frac{[IS]_{ki} [SI]_{ij}}{[S]_i} = \frac{[IS]_{ki} [SI]_{ij}}{N_i - ([IS]_{ip} + [II]_{ip})/a_{ip}}; \quad (4c)$$

$$[ISI]_{ijk}' = \frac{[IS]_{ij} [IS]_{kj}}{[S]_j} = \frac{[IS]_{ij} [IS]_{kj}}{N_j - ([IS]_{jq} + [II]_{jq})/a_{jq}}. \quad (4d)$$

In (4a), every susceptible  $G_i$ -node on average has  $[SS]_{ij}/[S]_i$  edges pointing to  $G_j$ -nodes.  $[S]_i$  is then estimated by employing (2e), where  $p$  and  $q$  are introduced to solve the equation. Here,  $p$  satisfies  $[IS]_{ip} + [II]_{ip} > 0$  and  $a_{ip} > 0$ ;  $q$  satisfies  $[IS]_{jq} + [II]_{jq} > 0$  and  $a_{jq} > 0$ . (4b)–(4d) can be similarly obtained.

By substituting (4a)–(4d) into (3a)–(3f), (3c)–(3f) can be rewritten as

$$\begin{aligned} \frac{d[SS]_{ij}}{dt} = & \gamma [SI]_{ij} + \gamma [IS]_{ij} \\ & - \sum_k \beta \frac{[IS]_{ki} [SS]_{ij}}{N_i - ([IS]_{ip} + [II]_{ip})/a_{ip}} \\ & - \sum_k \beta \frac{[IS]_{kj} [SS]_{ij}}{[S]_j}; \end{aligned} \quad (5a)$$

$$\begin{aligned} \frac{d[SI]_{ij}}{dt} = & -\gamma [SI]_{ij} + \gamma [II]_{ij} \\ & + \sum_k \beta \frac{[IS]_{kj} [SS]_{ij}}{N_j - ([IS]_{jq} + [II]_{jq})/a_{jq}} \\ & - \sum_k \beta \frac{[IS]_{ki} [SI]_{ij}}{N_i - ([IS]_{ip} + [II]_{ip})/a_{ip}}; \end{aligned} \quad (5b)$$

$$\begin{aligned} \frac{d[IS]_{ij}}{dt} = & -\gamma [IS]_{ij} + \gamma [II]_{ij} - \beta [IS]_{ij} \\ & + \sum_k \beta \frac{[IS]_{ki} [SS]_{ij}}{N_i - ([IS]_{ip} + [II]_{ip})/a_{ip}} \\ & - \sum_k \beta \frac{[IS]_{ij} [IS]_{kj}}{N_j - ([IS]_{jq} + [II]_{jq})/a_{jq}}; \end{aligned} \quad (5c)$$

$$\begin{aligned} \frac{d[II]_{ij}}{dt} = & -2\gamma [II]_{ij} + \beta [IS]_{ij} \\ & + \sum_k \beta \frac{[IS]_{ki} [SI]_{ij}}{N_i - ([IS]_{ip} + [II]_{ip})/a_{ip}} \\ & + \sum_k \beta \frac{[IS]_{ij} [IS]_{kj}}{N_j - ([IS]_{jq} + [II]_{jq})/a_{jq}}. \end{aligned} \quad (5d)$$

The terms of  $[SS]_{ij}$  can be suppressed by substituting (2b) into (5a)–(5d). As a result, we have

$$\begin{aligned} \frac{d[SI]_{ij}}{dt} = & -\gamma [SI]_{ij} + \gamma [II]_{ij} \\ & + \sum_k \beta \frac{[IS]_{kj} (a_{ij} N_i - [SI]_{ij} - [IS]_{ij} - [II]_{ij})}{N_j - (1/a_{jq}) ([IS]_{jq} + [II]_{jq})} \\ & - \sum_k \beta \frac{[IS]_{ki} [SI]_{ij}}{N_i - (1/a_{ip}) ([IS]_{ip} + [II]_{ip})}; \end{aligned} \quad (6a)$$

$$\begin{aligned}
& \frac{d[IS]_{ij}}{dt} \\
&= -\gamma [IS]_{ij} + \gamma [II]_{ij} - \beta_{ij} [IS]_{ij} \\
&+ \sum_k \beta \frac{[IS]_{ki} (a_{ij} N_i - [SI]_{ij} - [IS]_{ij} - [II]_{ij})}{N_i - (1/a_{ip}) ([IS]_{ip} + [II]_{ip})} \\
&- \sum_k \beta \frac{[IS]_{ij} [IS]_{kj}}{N_j - (1/a_{jq}) ([IS]_{jq} + [II]_{jq})}; \quad (6b)
\end{aligned}$$

$$\begin{aligned}
& \frac{d[II]_{ij}}{dt} \\
&= -2\gamma [II]_{ij} + \beta [IS]_{ij} \\
&+ \sum_k \beta \frac{[IS]_{ki} [SI]_{ij}}{N_i - (1/a_{ip}) ([IS]_{ip} + [II]_{ip})} \\
&+ \sum_k \beta \frac{[IS]_{ij} [IS]_{kj}}{N_j - (1/a_{jq}) ([IS]_{jq} + [II]_{jq})}. \quad (6c)
\end{aligned}$$

Near the disease-free equilibrium point, (6a)–(6c) can be linearized by suppressing all higher order terms. As a result, we have

$$\frac{d[SI]_{ij}}{dt} \approx -\gamma [SI]_{ij} + \gamma [II]_{ij} + \sum_k \frac{\beta a_{ij} N_i [IS]_{kj}}{N_j}; \quad (7a)$$

$$\frac{d[IS]_{ij}}{dt} \approx -(\gamma + \beta) [IS]_{ij} + \gamma [II]_{ij} + \sum_k \beta a_{ij} [IS]_{ki}; \quad (7b)$$

$$\frac{d[II]_{ij}}{dt} \approx -2\gamma [II]_{ij} + \beta [IS]_{ij}. \quad (7c)$$

After the model has been linearized, the condition of the disease-free equilibrium point can be obtained by performing eigenvalue analysis on the Jacobian matrix of the linearization. The Jacobian matrix is a  $3n^2 \times 3n^2$  matrix and denoted by  $\mathbf{J}$ . The nonlinear dynamic system is stable at the equilibrium point if and only if all the eigenvalues of the Jacobian matrix are negative, as stated by the Hartman-Grobman Theorem [32]. In other words, the epidemic is certainly extinct if

$$\lambda_1(\mathbf{J}) < 0, \quad (8)$$

where  $\lambda_1(\cdot)$  is the largest eigenvalue of the operator.  $\mathbf{J}$  is the Jacobian matrix of (7a)–(7c) and can be given by

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}^{11} & \mathbf{J}^{12} & \mathbf{J}^{13} \\ \mathbf{J}^{21} & \mathbf{J}^{22} & \mathbf{J}^{23} \\ \mathbf{J}^{31} & \mathbf{J}^{32} & \mathbf{J}^{33} \end{bmatrix} = \begin{bmatrix} \mathbf{J}^{11} & \mathbf{J}^{12} & \mathbf{J}^{13} \\ \mathbf{0} & \mathbf{J}^* & \end{bmatrix}, \quad (9)$$

where  $\mathbf{J}^{11}$  is an  $n^2 \times n^2$  diagonal matrix whose diagonal entries are  $-\gamma$ . We note that  $\mathbf{J}^{21}$  and  $\mathbf{J}^{31}$  are zero matrices. As a result,  $\mathbf{J}$  is an upper block triangular matrix, and

$$\lambda(\mathbf{J}) = \lambda(\mathbf{J}^{11}) \cup \lambda(\mathbf{J}^*), \quad (10)$$

where  $\lambda(\cdot)$  is the set of eigenvalues of the operator.

Here,  $\mathbf{J}^{23}$ ,  $\mathbf{J}^{32}$ , and  $\mathbf{J}^{33}$  are  $n^2 \times n^2$  diagonal matrices. Their diagonal entries are  $\gamma$ ,  $\beta$ , and  $-2\gamma$ , respectively. The entry in the  $(n(i-1)+j)$ -th row,  $(n(k-1)+l)$ -th column of  $\mathbf{J}^{22}$ , denoted by  $J_{ij,kl}^{22}$ , is given by

$$J_{ij,kl}^{22} = \begin{cases} a_{ij}\beta - (\gamma + \beta), & \text{if } ij = kl, i = l \\ -(\gamma + \beta), & \text{if } ij = kl, i \neq l \\ a_{ij}\beta, & \text{if } ij \neq kl, i = l \\ 0, & \text{if } ij \neq kl, i \neq l. \end{cases} \quad (11)$$

By employing the Schur complement, i.e.,

$$\begin{bmatrix} \mathbf{J}^{22} & \mathbf{J}^{23} \\ \mathbf{J}^{32} & \mathbf{J}^{33} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -(\mathbf{J}^{33})^{-1} \mathbf{J}^{32} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{J}^{23} \\ \mathbf{0} & \mathbf{J}^{33} \end{bmatrix}, \quad (12)$$

we have  $\lambda(\mathbf{J}^*) = \lambda(\mathbf{J}^{33}) \cup \lambda(\mathbf{H})$ , where  $\mathbf{H} = \mathbf{J}^{22} - \mathbf{J}^{23} \mathbf{J}^{33}^{-1} \mathbf{J}^{32}$ . The entry in the  $(n(i-1)+j)$ -th row,  $(n(k-1)+l)$ -th column of  $\mathbf{H}$ , denoted by  $H_{ij,kl}$ , is given by

$$H_{ij,kl} = \begin{cases} a_{ij}\beta - \left(\gamma + \frac{\beta}{2}\right), & \text{if } ij = kl, i = l \\ -\left(\gamma + \frac{\beta}{2}\right), & \text{if } ij = kl, i \neq l \\ a_{ij}\beta, & \text{if } ij \neq kl, i = l \\ 0, & \text{if } ij \neq kl, i \neq l. \end{cases} \quad (13)$$

We conclude that  $\lambda(\mathbf{J}) = \lambda(\mathbf{J}^{11}) \cup \lambda(\mathbf{J}^{33}) \cup \lambda(\mathbf{H})$ . Note that  $\mathbf{J}^{11}$  and  $\mathbf{J}^{33}$  are diagonal matrices, and all the diagonal entries of them are negative, i.e., all the eigenvalues of  $\mathbf{J}^{11}$  and  $\mathbf{J}^{33}$  are negative. As a result, we have

$$\lambda_1(\mathbf{J}) < 0 \iff \lambda_1(\mathbf{H}) < 0. \quad (14)$$

The matrix  $\mathbf{H}$  can be written as

$$\mathbf{H} = \mathbf{P} + \mathbf{Q}, \quad (15)$$

where  $\mathbf{Q}$  is a diagonal matrix and can be given by  $\mathbf{Q} = \text{diag}[-(\gamma + \beta/2)]$ . The entry in the  $(n(i-1)+j)$ -th row,  $(n(k-1)+l)$ -th column of  $\mathbf{P}$ , denoted by  $P_{ij,kl}$ , is given by

$$P_{ij,kl} = \begin{cases} a_{ij}\beta, & \text{if } i = l \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

We have that all the eigenvalues of  $\mathbf{Q}$  are  $-(\gamma + \beta/2)$  and

$$\lambda(\mathbf{H}) = \lambda(\mathbf{P}) - \left(\gamma + \frac{\beta}{2}\right). \quad (17)$$

Note that  $\mathbf{P}$  is a  $n^2 \times n^2$  sparse matrix and similar with a block matrix consisting of  $\beta \mathcal{A}$  and zero matrices as given by

$$\mathbf{P} = \mathbf{X}^{-1} \begin{bmatrix} \beta \mathcal{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{X}. \quad (18)$$

This is achieved by performing matrix operations on  $\mathbf{P}$ . Thus, the eigenvalues of  $\mathbf{P}$  can be given by

$$\lambda(\mathbf{P}) = \beta \times \lambda(\mathcal{A}) \cup 0. \quad (19)$$

Combining (8), (14), (17), and (19), the epidemic will become extinct if  $\beta\lambda_1(\mathcal{A}) - (\gamma + \beta/2) < 0$ . In other words, the epidemic threshold, denoted by  $\tau^*$ , is given by

$$\tau^* = \frac{1}{\lambda_1(\mathcal{A}) - 1/2}. \quad (20)$$

The epidemic dies out, if  $\tau = \beta/\gamma < \tau^*$ .

## 5. Simulation and Numerical Results

In this section, numerical and simulation results are presented to validate the proposed group-based SIS model and the deduced epidemic threshold. In every run of the simulations, the groups and the network topology, i.e., the scale and structure, are first set up, according to the rules specified in Section 3; i.e.,  $a_{ij}$  number of directed edges are added from a  $G_i$ -node to  $a_{ij}$  number of randomly selected  $G_j$ -nodes. The nodes do not connect themselves despite  $a_{ii} \geq 0$ . Then, the infection rate  $\beta$  and the curing rate  $\gamma$  are configured, based on the analytical epidemic threshold  $\tau^*$ . The simulations are carried out on the NepidemiX [33], which is a Python library implementing simulations of epidemics. For initialization, randomly selected 10% of the nodes are infected. During a simulation run, the infected nodes can be cured at the rate of  $\gamma$ . Every infected node can infect its neighbors connected by edges at the rate of  $\beta$ . Every dot in the figures is the average of 100 independent runs under the same configurations, including the network topology, the percentage of initially infected nodes,  $\beta$  and  $\gamma$ .

We first validate our model on a 1000-node network where nodes connect each other and form a complete graph. The infection and curing rates are set to be  $\beta = 0.0005$  and  $\gamma = 0.1$  per time slot.  $\tau = \beta/\gamma = 0.005$  is set to be larger than  $\tau^* = 0.001$  to evaluate the proposed model. Ten percent of nodes are randomly chosen to be infected at  $t = 0$ . Figure 3 plots the infection density, given by  $\sum [I]_i/N$ , from  $t = 0$  to 30. The analytical results are numerically evaluated by employing (6a)–(6c) with the nodes evenly divided into 1, 2, 5, 10, 25, and 50 groups, respectively.

From Figure 3, we can see that the analytical results can asymptotically approach the simulation results with the increasing number of groups, e.g., from  $n = 1$  to  $n = 50$ . The simulation results can outgrow the analytical results when the number of node groups is small. The analytical results under a single group of nodes, i.e.,  $n = 1$ , can substantially underestimate the infection density. The analytical result is only 83.8% (0.67/0.8) of the simulation result, when  $t = 30$ . In contrast, the analytical results under 50 groups of nodes, i.e., when  $n = 50$ , match the simulation results indistinguishably. This is because the proposed model is designed to decouple the state transitions of edges connecting nodes from different groups and estimate the number of edges connecting infected nodes and subsequently the population of infected nodes. The

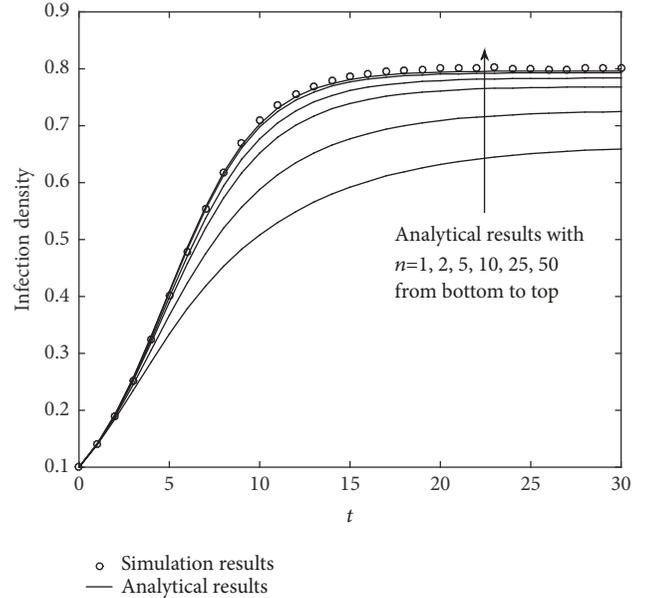


FIGURE 3: The growth of infection density where a complete graph with 1000 nodes is considered.  $\beta = 0.0005$  and  $\gamma = 0.1$ . The analytical results are obtained based on (6a)–(6c) by evenly dividing the nodes into 1, 2, 5, 10, 25, and 50 groups.

mean-field approximation is applied to model the interplay between the averaged ratios of infectious edges connecting different pairs of node groups. With an increasing number of node groups, the averaged ratio of infectious edges connecting a specific pair of node groups can become increasingly representative with a reducing deviation. In other words, the averaged ratio becomes increasingly precise for a reducing set of edges. In the special case where every node forms a group (i.e., the number of groups is  $n = 1000$ ), the ratio is exactly the probability at which an edge is infectious. The proposed model is able to capture the state transition of an edge under the averaged effect of all other individual edges, and can be fairly accurate given the large number of edges. Note that the number of groups, e.g.,  $n = 50$ , is far less than the network size, i.e., 1000. This finding allows us to model the epidemic propagation with a small number of differential equations. Figure 3 also shows that the analytical results can be accurate at the initial stage (or low infection densities) even with few node groups. This is because the state transitions of different edges are loosely coupled if only very few nodes are infected.

We proceed to evaluate the model accuracy with different infection densities. This is done by adjusting the infection rate. A 1000-node network is considered where nodes connect each other and form a complete graph. The epidemic propagation with four effective spreading rates, i.e.,  $\tau = \beta/\gamma = 0.002, 0.0015, 0.0011, \text{ and } 0.001$ , are simulated and analyzed. The values of  $\tau$  are larger than, and close to, the analytical threshold. The infection rates are obtained by adjusting  $\beta$  while setting  $\gamma$  to 0.1. The analysis is based on (6a)–(6c) by evenly dividing the nodes into 5 groups to explore the applicability of the model to a small number of groups.

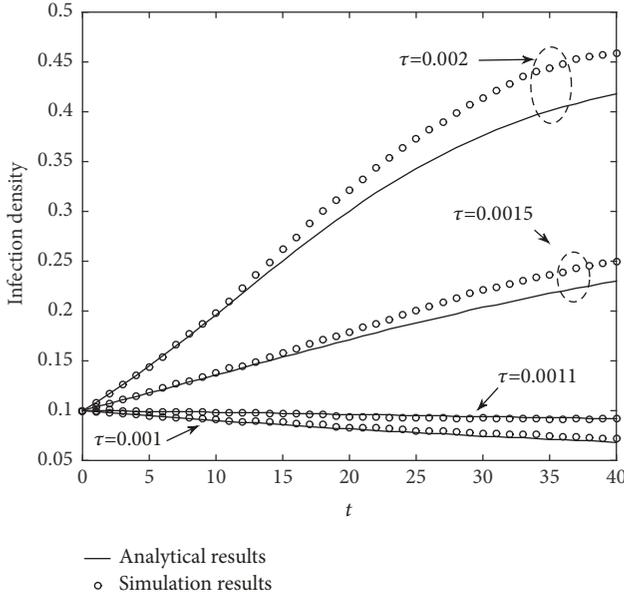


FIGURE 4: The infection density with the growth of time where a complete graph with 1000 nodes is considered. The analytical results are obtained based on (6a)–(6c) by evenly dividing the nodes into 5 groups where  $\tau = \beta/\gamma = 0.002, 0.0015, 0.0011$ , and  $0.001$ , respectively.

From Figure 4, we can see that the simulation results still overtake the analytical results when  $\tau = 0.0015$  and  $0.002$ . For example, the simulation result is  $0.459$  in the case of  $t = 40$  and  $\tau = 0.002$ , while the analytical result is only  $0.418$ . However, the gap between simulation results and analytical results decreases with dropping  $\tau$ , i.e., from  $\tau = 0.002$  to  $0.0015$ . When  $\tau$  further declines, i.e.,  $\tau = 0.0011$  and  $0.001$ , the analytical results are able to match the simulation results from the beginning to the end. According to (20), the epidemic threshold is given by  $\tau^* = 0.001$ . This figure reveals that the proposed mean-field model is asymptotically accurate with a decreasing  $\tau$  and is able to precisely describe the epidemic propagation when the effective spreading rate is around the epidemic threshold.

We evaluate the epidemic threshold  $\tau^*$  given by (20) in Figure 5. Three networks with 500 nodes are considered, where nodes are divided into three groups (i.e.,  $\mathcal{N} = [100, 200, 200]$ ). The impact of the network topology on the threshold is evaluated by varying the number of edges in the network. Without loss of generality, their collapsed adjacency matrix  $\mathcal{A}$  is set to

$$\mathcal{A} = \alpha \begin{bmatrix} 10 & 20 & 20 \\ 10 & 20 & 20 \\ 10 & 20 & 20 \end{bmatrix}, \quad (21)$$

where  $\alpha = 4, 6$ , or  $8$ . The curing rate  $\gamma$  is set to be  $0.1$ . Ten percent of nodes are randomly selected to be infected at the initial state. The infection density at  $t = 1000$  is used as the stable infection density, as indicated by the  $y$ -axis. Evaluated with (20), the epidemic threshold is  $\tau^* = 0.005, 0.0033$ , and  $0.0025$  when  $\alpha = 4, 6$ , and  $8$ , respectively. We can see from

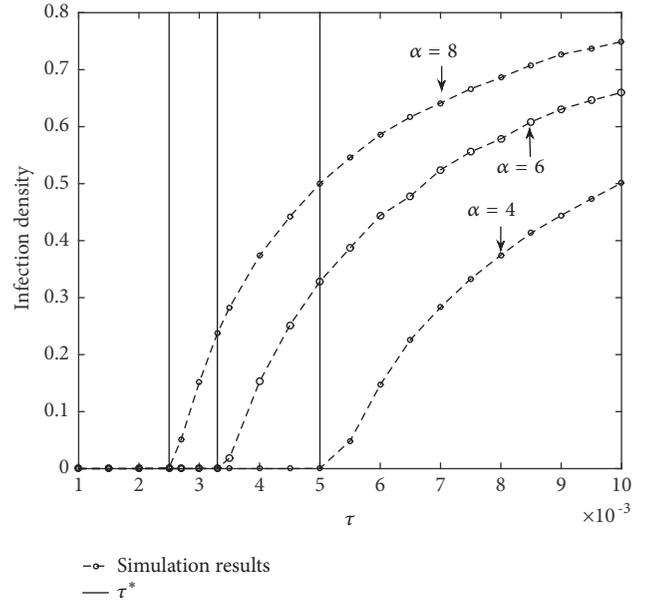


FIGURE 5: The validation of epidemic threshold given by (20), where the  $y$ -axis is the infection density at  $t = 1000$ . Three networks with 500 nodes are considered.  $\gamma$  is set to be  $0.1$ .  $\alpha = 4, 6, 8$  is used to adjust the number of edges.

Figure 5 that  $\tau^*$  (the solid vertical lines) can precisely specify the epidemic thresholds. When  $\tau < \tau^*$ , the epidemic can be suppressed eventually. When  $\tau > \tau^*$ , the infection density grows with  $\tau$  and also exhibits convexity. We can see that the network topology has a strong impact on the epidemic threshold and the infection density. Specifically, the threshold decreases with the growth of  $\alpha$ . For example,  $\tau$  halves from  $0.005$  to  $0.0025$ , when  $\alpha$  doubles from  $4$  to  $8$  (the number of edges doubles, as well). The infection density increases with the growth of  $\alpha$ , especially around the threshold. For example, in the case of  $\alpha = 8$  and  $\tau = 0.005$ , the infection density is  $0.5$  as compared to the infection density of  $\alpha = 4$ .

We note that the epidemic threshold, given by (20), is determined by the network structure  $\mathcal{A}$ , rather than the number of nodes given by  $\mathcal{N}$ . To illustrate this, we compare the number of infected populations in different scales of networks, illustrated by Figure 6. To be specific,  $\mathcal{N} = \delta \times [10, 20, 20]$ , where  $\delta = 2, 4$  and  $6$ , respectively. Their collapsed adjacency matrices are obtained with (21) by letting  $\alpha = 1$ . As a result,  $\tau^* = 0.02$ . Figure 6 firstly confirms the accuracy of  $\tau^*$  illustrated by the solid vertical line. The observation that the three networks with the same structure but different scales share the same epidemic threshold validates that the epidemic threshold depends on the network structure rather than the scale. This finding can significantly reduce the complexity to deduce the epidemic threshold, i.e., from decomposing an  $N$ -dimensional matrix by using  $1/\lambda_1(\mathbf{A})$ , given by [12], to decomposing an  $n$  dimensional matrix ( $n \ll N$ ). For example,  $N = 1000$  and  $n = 5$  in Figure 4, and  $N = 500$  and  $n = 3$  in Figure 5. It is interesting to notice that the infection densities are the same, e.g.,  $79.6/100 \approx 158.8/200 \approx 240/300$  in the case of  $\tau = 0.1$ ,

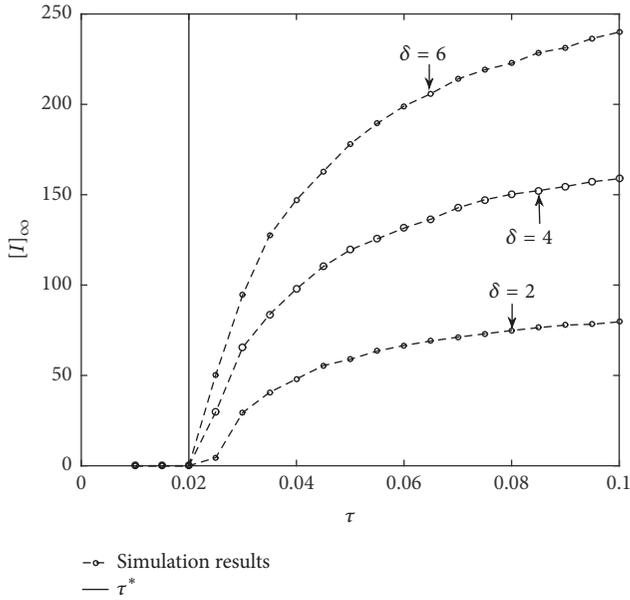


FIGURE 6: The validation of epidemic threshold, where the  $y$ -axis is the infected population at  $t = 1000$ . Different scales of networks ( $N=100, 200, \text{ and } 300$ , respectively) are considered.  $\gamma$  is set to be  $0.1$ .

although the infected population varies in different scales of networks. This can be the reason of that the epidemic threshold depends on the network structure rather than the network scale.

## 6. Conclusion

In this paper, we designed a continuous-time SIS model in large-scale networks. By categorizing nodes into groups, the model complexity was significantly reduced. The proposed epidemic model was validated to be asymptotically accurate with the decrease of the effective spreading rate and/or the increase of node groups. The epidemic threshold can be deduced with the largest eigenvalue of the collapsed adjacency matrix whose dimension is much smaller than the network scale. Simulation results corroborated the effectiveness of the model, as well as the analytical accuracy of the threshold in large-scale networks.

## Data Availability

The simulations are built on the epidemic simulation platform NepidemiX.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Key R&D Program of China (no. 2017YFB0802703), the National Natural Science

Foundation of China (no. 61121061), and UTS DVC-R Funding Initiative for Research Strengths.

## References

- [1] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 343–359, IEEE, Oakland, CA, USA, May 1991.
- [2] X. Wang, K. Zheng, X. Niu, B. Wu, and C. Wu, "Detection of command and control in advanced persistent threat based on independent access," in *Proceedings of the 2016 IEEE International Conference on Communications, ICC 2016*, pp. 1–6, IEEE, Kuala Lumpur, Malaysia, May 2016.
- [3] W. Yang, Y. Qin, and Y. Yang, "Analysis of malicious flows via SIS epidemic model in CCN," in *Proceedings of the IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*, pp. 748–753, IEEE, Honolulu, HI, USA, April 2018.
- [4] Z. Kong and E. M. Yeh, "Resilience to degree-dependent and cascading node failures in random geometric networks," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 56, no. 11, pp. 5533–5546, 2010.
- [5] R. M. Anderson, *Infectious Diseases of Humans: Dynamics And Control*, vol. 28, Wiley Online Library, 1992.
- [6] X. Wang, X. Zha, G. Yu et al., "Attack and defence of ethereum remote apis," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps'18)*, IEEE, Abu Dhabi, UAE, 2018.
- [7] X. Wang, W. Ni, K. Zheng, R. P. Liu, and X. Niu, "Virus Propagation Modeling and Convergence Analysis in Large-Scale Networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2241–2254, 2016.
- [8] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 1–14, 2009.
- [9] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: an eigenvalue viewpoint," in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems (SRDS '03)*, pp. 25–34, IEEE, Florence, Italy, October 2003.
- [10] L. P. Kadanoff, "More is the same; phase transitions and mean field theories," *Journal of Statistical Physics*, vol. 137, no. 5, pp. 777–797, 2009.
- [11] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proceedings of the Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 2, pp. 1455–1466, IEEE, Miami, FL, USA, 2005.
- [12] P. Van Mieghem and R. Van De Bovenkamp, "Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks," *Physical Review Letters*, vol. 110, no. 10, Article ID 108701, 2013.
- [13] A. Barabási and E. Bonabeau, "Scale-Free Networks," *Scientific American*, vol. 288, no. 5, pp. 60–69, 2003.
- [14] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, no. 3, pp. 925–979, 2015.
- [15] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

- [16] H. Kang and X. Fu, "Epidemic spreading and global stability of an SIS model with an infective vector on complex networks," *Communications in Nonlinear Science & Numerical Simulation*, vol. 27, no. 1–3, pp. 30–39, 2015.
- [17] T. Li, X. Liu, J. Wu, C. Wan, Z.-H. Guan, and Y. Wang, "An epidemic spreading model on adaptive scale-free networks with feedback mechanism," *Physica A: Statistical Mechanics and its Applications*, vol. 450, pp. 649–656, 2016.
- [18] J. O. Kephart and S. R. White, "Measuring and modeling computer virus prevalence," in *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 2–15, IEEE, Oakland, Calif, USA, May 1993.
- [19] M. Vojnovic and A. J. Ganesh, "On the race of worms, alerts, and patches," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1066–1079, 2008.
- [20] Y. Yao, Q. Fu, W. Yang, Y. Wang, and C. Sheng, "An epidemic model of computer worms with time delay and variable infection rate," *Security and Communication Networks*, vol. 2018, Article ID 9756982, 11 pages, 2018.
- [21] L. Liu, R. K. L. Ko, G. Ren, and X. Xu, "Malware propagation and prevention model for time-varying community networks within software defined networks," *Security and Communication Networks*, vol. 2017, Article ID 2910310, 8 pages, 2017.
- [22] C. C. Zou, D. Towsley, and W. B. Gong, "Modeling and simulation study of the propagation and defense of internet e-mail worms," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 2, pp. 105–118, 2007.
- [23] C. Li, R. van de Bovenkamp, and P. Van Mieghem, "Susceptible-infected-susceptible model: A comparison of n-intertwined and heterogeneous mean-field approximations," *Physical Review E*, vol. 86, no. 3, p. 026116, 2012.
- [24] W. K. Chai and G. Pavlou, "Path-Based Epidemic Spreading in Networks," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 565–578, 2017.
- [25] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *International Journal of Information Management*, vol. 38, no. 1, pp. 86–96, 2018.
- [26] L. A. Meyers, M. E. Newman, and B. Pourbohloul, "Predicting epidemics on directed contact networks," *Journal of Theoretical Biology*, vol. 240, no. 3, pp. 400–418, 2006.
- [27] C. Li, H. Wang, and P. Van Mieghem, "Epidemic threshold in directed networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 88, no. 6, p. 062802, 2013.
- [28] A. Khanafer, T. Basar, and B. Ghahsifard, "Stability of epidemic models over directed graphs: a positive systems approach," *Automatica*, vol. 74, pp. 126–134, 2016.
- [29] A. Lyapunov, "The general problem of the stability of motion," *International Journal of Control*, vol. 55, no. 3, pp. 531–534, 1992, English Translation.
- [30] M. J. Keeling, "The effects of local spatial structure on epidemiological invasions," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 266, no. 1421, pp. 859–867, 1999.
- [31] K. J. Sharkey, C. Fernandez, K. L. Morgan et al., "Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks," *Journal of Mathematical Biology*, vol. 53, no. 1, pp. 61–85, 2006.
- [32] D. K. Arrowsmith and C. M. Place, *Dynamical Systems: Differential Equations, Maps, and Chaotic Behaviour*, Chapman & Hall, London, UK, 1992.
- [33] L. Ahrenberg, "Nepidemix," <http://nepidemix.irmacs.sfu.ca>.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

