Modelling multi-state diffusion process in complex networks: theory and applications

Yishi Lin[†] and John C. S. Lui

Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong
[†]Corresponding author. Email: yslin@cse.cuhk.edu.hk

Kyomin Jung

Department of Electrical & Computer Engineering, Seoul National University, Seoul, South Korea

and

Sungsu Lim

Department of Mathematical Sciences, KAIST, Daejeon, South Korea

Edited by: Hocine Cherifi

Recent years have witnessed a growing interest in understanding the fundamental principles of how epidemic, ideas or behaviour spread over large networks (e.g. the Internet or online social networks). The conventional approach is to use the susceptible-infected-susceptible (SIS) model or its derivatives. We like to note that these models are often too restrictive and may not be applicable in many realistic situations. In this paper, we propose a 'generalized SIS model' by allowing the existence of intermediate states between susceptible and infected states. To analyse the diffusion process of the generalized SIS model on large graphs, we use the 'mean-field analysis technique' to determine which initial condition leads to or prevents the outbreak of information or virus. For any general connected graphs, we show that the condition which can prevent the spread of contagions depends on two de-coupled effects: the network topology and the parametric values of the generalized SIS model. Experimental results based on both synthetic and real-world datasets show that our methodology can accurately predict the behaviour of the phase-transition process for any general graphs. We also extend our generalized SIS model to analyse the dynamics and behaviour of *two competing sources*. This is useful if one wants to model competing products in a large network or competition between virus and antidote in a large communication network. We present the analytical derivation and show via experiment how different factors such as initial condition, transmission rates, recovery rates or the number of states can affect the phase transition process and the final equilibrium. Our models and methodology can serve as an essential tool in analysing and understanding the information diffusion process in large networks.

Keywords: models of complex networks; applications of complex network analysis; complex networks and epidemics; generalized SIS model; tipping point; competing sources.

1. Introduction

In the area of complex networks or network science, many researchers have studied the problem of contagion [1–3]: How viruses, ideas or behaviours can be spread in large networks. Understanding the dynamical process of contagion is of importance if one wants to prevent and control the spread of

diseases, or to maximize the influence of a product in online social networks [4]. One popular contagion model is the susceptible-infected-susceptible (SIS) model. Under this model, each node in a network can be in one of these two states: 'susceptible' (*S*) or 'infected' (*I*). When a node is in state *S*, it is subjected to be influenced by its neighbouring nodes in the *I* state, while only nodes in the *I* state can influence their neighbouring nodes. When a node is in the 'infected' state, it becomes recovered and returns back to the 'susceptible" state at some rate. Hence, using the SIS model, one can model the spread of contagions such as flu or ideas in a large network.

We like to note that the SIS model is often too restrictive. For example, consider the case where a diffusion model is used to describe a product adoption [5]. Some time after the product release, some people may have purchased the product while some may not. The recent results of consumer purchase decision process theory [6] reveal that, in general, there are five stages until a consumer makes a purchase and influences others. These states include 'product recognition', 'information search', 'alternative evaluation', 'purchase decision' and 'post-purchase behaviour'. This implies that using the traditional SIS model of having a single state (e.g., *susceptible* state) to model the purchase decision process is insufficient, and one needs to further expand the *susceptible* state into more states according to the degree of interest. Moreover, in some scenarios, there are more than one contagion, ideas or behaviours spreading in the network. We also need to consider the competition among multiple sources so that we can study how the interaction may affect the final state of the network.

Contributions: From these aspects, we make several contributions in this paper. They are as follows:

- Our first contribution is that we propose a generalization of the SIS model by allowing the number of states *k* to adoption (or infection) be > 1. That is, the states can be from state 0 to state *k* − 1, where the state *k* − 1 is the *infected* (*active*) state: An infected node can influence its neighbouring nodes, while nodes in state 0 to *k* − 1 can be promoted to a higher state if they are exposed to its infected neighbour. We use the *multidimensional mean-field method* to analyse the influence spreading dynamics in complete and general graphs, and determine the condition of phase transition. We show that the condition that prevents the spread of contagions depends on two de-coupled effects: the network topology and the parametric values of the generalized SIS model. For undirected graphs, our results show that this condition depends on the largest eigenvalue of the adjacency matrix, the infection rates, the recovery rate and the number of states *k* in the generalized SIS model. For directed graphs, it also depends on the maximum in-degree among all nodes besides the above factors.

- Our second contribution is to model and analyse the dynamics and behaviour of *two competing sources*. We first use our generalized SIS model in two competing sources, with one being dominant and the other being recessive, and these two sources compete with each other in the network. We formulate the dynamic process and show how different factors, such as different initial conditions and transmission rates may affect the phase transition and the final equilibrium.

- Last but not least, our methodology can predict the behaviour of the diffusion accurately, and we use several applications to illustrate how one can design simple and effective vaccination or advertisement strategy in a large network.

The rest of the paper is organized as follows: We first propose and analyse the generalized SIS model under various network settings in Section 2. Experimental results on the effectiveness and accuracy of our analytical framework in Section 3. Then we present the generalized SIS model with two competing sources in Section 4, and show the corresponding experimental results in Section 5. Related work is in given Section 6 and finally we conclude our work in Section 7.
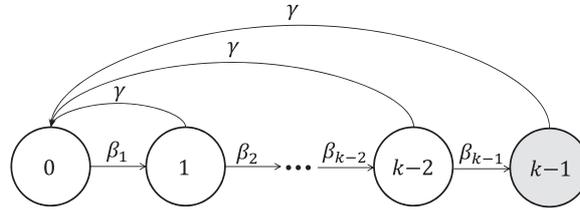
FIG. 1. The generalized SIS model with $k \geqslant 2$ states.

2. Generalized SIS model and analysis

Let us first formally state our multi-state SIS model. We first model the network under study as a fully connected undirected graph $G = (V, E)$ (later, we will generalize our results to general graphs). Any node $v \in V$ can be in one of $k \geqslant 2$ states: $\{0, 1, \ldots, k-1\}$. Only nodes in state $k-1$ (which we call the infected or active state) can increase the state value of its neighbours, say node s, from state $j \in \{0, 1, \ldots, k-2\}$ to state $j+1$ with an infection rate of $\beta_{j+1} > 0$. Each node can recover with a recovery rate of $\gamma > 0$ and its state will return to state 0. Figure 1 depicts our generalized SIS model with $k \geqslant 2$ states.

Let us first briefly review some work on the analysis of the classical SIS (or $k = 2$) model. Under the classical SIS model, state 0 corresponds to *susceptible* (S), while state 1 corresponds to *infected* (I). The infection rate is β and the recovery rate is γ. Let $x_0(t)$ and $x_1(t)$ be the fraction of nodes in state S and state I at time $t \geqslant 0$, respectively. We define (x_0, x_1) as an equilibrium for the model, i.e. $\lim_{t \to \infty} x_i(t)$, $i = 1, 2$. If G is a fully connected graph, we have $dx_1/dt = \beta x_0 x_1 - \gamma x_1$ and $x_0(t) + x_1(t) = 1$. For this model, we have two possible equilibria: one is $(x_0, x_1) = (1, 0)$ and the other is $(x_0, x_1) = (\gamma/\beta, 1 - \gamma/\beta)$.

Let us consider an arbitrary graph G. Denote A as the adjacency matrix of G, i.e. $A_{ij} = 1$ if node i and j are neighbours to each other, and 0 otherwise. Let $x_0^{(i)}(t)$ and $x_1^{(i)}(t)$ be the probability of node i in state S and state I at time t, where $i \in V$, respectively. Denote $(x_0^{(i)}, x_1^{(i)})_{i \in V}$ as an equilibrium for this model. We have $dx_1^{(i)}/dt = \beta x_0^{(i)} \sum_j A_{ji} x_1^{(j)} - \gamma x_1^{(i)} = 0$.

It is important for us to note that this is a 'mean-field approximation'. The right-hand side of the above equation contains three average quantities ($x_0^{(i)}(t)$, $x_1^{(i)}(t)$ and $x_1^{(j)}(t)$), and in multiplying these quantities, we implicitly assume that the product of their average is equal to the average of their product. For a large graph G, this form of mean-field approximation is very accurate. But for small networks, this may not hold since probabilities are not independent. We say that the infection process will die out when eventually all nodes in G are in state S, or $\lim_{t \to \infty} x_0^{(i)}(t) = 1$, $\forall i \in V$. One can formally show [7] that the condition for infection to die out over time is $\beta/\gamma < 1/\lambda_1$, and λ_1 is the largest eigenvalue of the adjacency matrix A.

2.1 *Ternary model for generalized SIS*

For the ease of presentation, we first consider a generalized SIS model with $k = 3$ states. State 0 is susceptible and state 1 means that a node is exposed but *not* infected yet. For each state $s \in \{0, 1, 2\}$, let $x_s(t)$ be the fraction of nodes in state s at time t and we let (x_0, x_1, x_2) be an equilibrium for the model. It is easy to see that $x_0(t) + x_1(t) + x_2(t) = 1$ for all t. Using the mean-field analysis, we have

the following system of differential equations which describes the system dynamics:

$$\frac{dx_2}{dt} = \beta_2 x_1 x_2 - \gamma x_2, \tag{2.1}$$

$$\frac{dx_1}{dt} = -\beta_2 x_1 x_2 + \beta_1 x_0 x_2 - \gamma x_1. \tag{2.2}$$

By setting $dx_2/dt = 0$, we have two cases to consider: $x_2 = 0$ or $x_2 \neq 0$ and $x_1 = \gamma/\beta_2$. The non-trivial equilibrium corresponds to the second case. By setting $dx_1/dt = 0$, we have $-\beta_2 x_1 x_2 + \beta_1 x_0 x_2 - \gamma x_1 = 0$, hence

$$x_0 = \left(\frac{\beta_2 x_2 + \gamma}{\beta_1 x_2} \right) x_1.$$

For this non-trivial solution, we have $x_2 \neq 0$, $x_1 = \gamma/\beta_2$ and $x_0 = ((\beta_2 x_2 + \gamma)/\beta_1 x_2)(\gamma/\beta_2)$. Because $\sum_{i=0}^{2} x_i = 1$, we have the following relationship: $1 = ((\beta_2 x_2 + \gamma)/\beta_1 x_2)(\gamma/\beta_2) + \gamma/\beta_2 + x_2$, or

$$\beta_1 \beta_2 x_2^2 + [\gamma(\beta_1 + \beta_2) - \beta_1 \beta_2]x_2 + \gamma^2 = 0. \tag{2.3}$$

Let us now give the necessary condition for the dynamic system (2.1) and (2.2) to reach a non-trivial equilibrium.

THEOREM 2.1 The necessary condition for the dynamic system as described in (2.1) and (2.2) to reach a non-trivial equilibrium is

$$\gamma \leqslant \frac{\beta_1 \beta_2}{(\sqrt{\beta_1} + \sqrt{\beta_2})^2}. \tag{2.4}$$

Proof. The discriminant of the quadratic Equation (2.3) is $D = [\gamma(\beta_1 + \beta_2) - \beta_1 \beta_2]^2 - 4\gamma^2 \beta_1 \beta_2$. Let us derive the conditions for $D \geqslant 0$. We could write $D = 0$ as a quadratic equation of γ. Then we can express $D = 0$ as

$$\begin{aligned} D &= \gamma^2 (\beta_1 + \beta_2)^2 - 2\gamma(\beta_1 + \beta_2)\beta_1 \beta_2 - 4\gamma^2 \beta_1 \beta_1 \\ &= \gamma^2 (\beta_1 - \beta_2)^2 - 2\gamma(\beta_1 + \beta_2)\beta_1 \beta_2 + (\beta_1 \beta_2)^2 = 0. \end{aligned}$$

We first consider the case when $\beta_1 \neq \beta_2$. The discriminant of the above quadratic equation of γ, denoted by D'_γ, is

$$\begin{aligned} D'_\gamma &= [2(\beta_1 + \beta_2)\beta_1 \beta_2]^2 - 4(\beta_1 - \beta_2)^2 (\beta_1 \beta_2)^2 \\ &= 4[(\beta_1 + \beta_2)^2 - (\beta_1 - \beta_2)^2](\beta_1 \beta_2)^2 \\ &= 16(\beta_1 \beta_2)^3. \end{aligned}$$
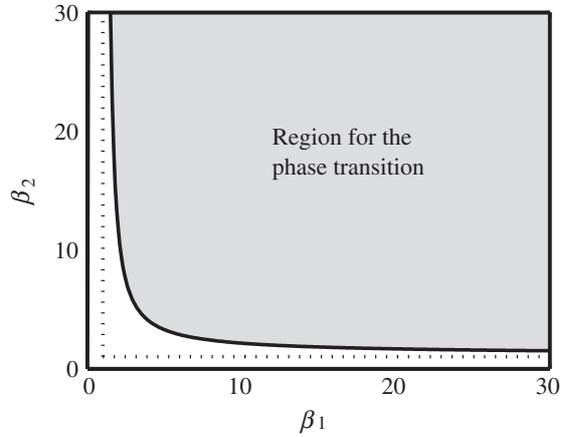
FIG. 2. The phase transition region of the ternary model where $\gamma = 1$.

We have $D'_\gamma > 0$ since we assume that the infection rates are positive. Thus, the equation $D = 0$ has two real solutions, and we denote them by γ_1 and γ_2.

$$\gamma_{1,2} = \frac{2(\beta_1 + \beta_2)\beta_1\beta_2 \pm 4\beta_1\beta_2\sqrt{\beta_1\beta_2}}{2(\beta_1 - \beta_2)^2}$$

$$= \frac{\beta_1\beta_2(\sqrt{\beta_1} \pm \sqrt{\beta_2})^2}{(\beta_1 - \beta_2)^2}$$

$$= \frac{\beta_1\beta_2}{(\sqrt{\beta_1} \pm \sqrt{\beta_2})^2}.$$

For $\beta_1 \neq \beta_2$, $D \geqslant 0$ holds if $\gamma \leqslant \beta_1\beta_2/(\sqrt{\beta_1} + \sqrt{\beta_2})^2$ or $\gamma \geqslant \beta_1\beta_2/(\sqrt{\beta_1} - \sqrt{\beta_2})^2$. Note that $x_1 = \gamma/\beta_2 \leqslant 1$ implies $\beta_2 \geqslant \gamma$. However, $\gamma \geqslant \beta_1\beta_2/(\sqrt{\beta_1} - \sqrt{\beta_2})^2 \geqslant \beta_2$ is a contradiction. Hence, for $\beta_1 \neq \beta_2$, the dynamic system has a non-trivial equilibrium which implies $\gamma \leqslant \beta_1\beta_2/(\sqrt{\beta_1} + \sqrt{\beta_2})^2$.

The other case we have to consider is when $\beta_1 = \beta_2$. $D \geqslant 0$ holds if and only if

$$\gamma \leqslant \frac{\beta_1\beta_2}{2(\beta_1 + \beta_2)} = \frac{\beta_1\beta_2}{(\sqrt{\beta_1} + \sqrt{\beta_2})^2}.$$

The last equality holds because $\beta_1 = \beta_2$. Hence we can conclude that the dynamic system has a non-trivial equilibrium which implies

$$\gamma \leqslant \frac{\beta_1\beta_2}{(\sqrt{\beta_1} + \sqrt{\beta_2})^2}. \qquad \square$$

Using this result, we can determine the *region for the phase transition*, or the necessary conditions, in which the majority of nodes are in state 2 at equilibrium. Figure 2 depicts the phase transition region. In this region, the information or virus has a chance to outbreak if (β_1, β_2) satisfies $\beta_1\beta_2/(\sqrt{\beta_1} + \sqrt{\beta_2})^2 \geqslant \gamma$. In Section 2.2, we extend this condition to a general case of $k \geqslant 2$.

Suppose that there exists at least one real solution for Equation (2.3) and $(x_1 = \gamma/\beta, x_2)$ is a non-trivial solution for the dynamic system. The solution is feasible if it satisfies $0 \leqslant x_1, x_2 \leqslant 1$ and $0 \leqslant x_1 + x_2 \leqslant 1$. For any feasible non-trivial solution, we are interested in whether the dynamic system will reach this equilibrium or not. Let us now present the stability condition of the ternary model.

THEOREM 2.2 For a feasible and non-trivial solution $(x_1 = \gamma/\beta_2, x_2)$ of the dynamic system as described in (2.1) and (2.2), it has a non-trivial equilibrium if $x_2 > \frac{1}{2}(1 - \gamma/\beta_1 - \gamma/\beta_2)$.

Proof. Define $f_1(x_1, x_2) = dx_1/dt = -\beta_2 x_1 x_2 + \beta_1(1 - x_1 - x_2)x_2 - \gamma x_1$ and $f_2(x_1, x_2) = dx_2/dt = \beta_2 x_1 x_2 - \gamma x_2$. We also define J as the Jacobian matrix where $J = (df_i/dx_j)_{i,j=1,2}$. A fixed point (or an equilibrium) of the system of differential equation which is defined by Equations (2.1) and (2.2) is stable if the trace (Tr) of J is negative and the determinant (Det) of J is positive. We can express the differential of f_i as follows:

$$\frac{df_1}{dx_1} = -\beta_2 x_2 - \beta_1 x_2 - \gamma, \quad \frac{df_1}{dx_2} = -\beta_2 x_1 + \beta_1 - \beta_1 x_1 - 2\beta_1 x_2,$$

$$\frac{df_2}{dx_1} = \beta_2 x_2, \quad \frac{df_2}{dx_2} = \beta_2 x_1 - \gamma.$$

Since $x_1 = \gamma/\beta_2$, the Jacobian matrix J is

$$J = \begin{bmatrix} -\beta_2 x_2 - \beta_1 x_2 - \gamma & -\gamma + \beta_1 - \dfrac{\beta_1 \gamma}{\beta_2} - 2\beta_1 x_2 \\ \beta_2 x_2 & 0 \end{bmatrix}.$$

For any feasible x_2, we have $Tr < 0$. Also, if $x_2 > \frac{1}{2}(1 - \gamma/\beta_1 - \gamma/\beta_2)$, we have $Det = -\beta_2 x_2(-\gamma + \beta_1 - \beta_1 \gamma/\beta_2 - 2\beta_1 x_2) > 0$, which completes the proof. □

Hence, a non-trivial equilibrium (x_1, x_2) is stable if $x_2 > \frac{1}{2}(1 - \gamma/\beta_1 - \gamma/\beta_2)$, or a saddle point elsewhere. Note that $(x_1, x_2) = (0, 0)$ is a stable equilibrium because the eigenvalues of J are negative.

EXAMPLE Consider the following. If $\beta_1 = \beta_2 = 20$ and $\gamma = 1$, then we have $x_1 = 0.05$ from $x_1 = \gamma/\beta_2$. From Equation (2.3), we have $400x_2 = (20x_2 + 1)^2$. Solving this, we obtain $x_2 = 0.003$ or 0.897. As $\frac{1}{2}(1 - \gamma/\beta_1 - \gamma/\beta_2) = 0.45 < 0.897$, we can conclude that $(x_1, x_2) = (0.05, 0.897)$ is a stable non-trivial equilibrium.

APPLICATION 1 (Vaccination strategy) We now illustrate how one can apply the results of this ternary SIS model. Assume that there is a large-scale computer virus spreading in a network; the detection rate of an anti-virus software γ needs to be high enough so as to prevent its spreading for a given (β_1, β_2). This leads to the following vaccination strategy: the value of γ needs to guarantee by the following conditions:

- Equations (2.1) and (2.2) have no solution, or

- Equation (2.3) has no feasible non-trivial solution of x_2, i.e. $x_2 \notin (0, 1]$, or

- Equation (2.3) has feasible solutions of x_2 but none of them are stable, i.e. $x_2 \leqslant \frac{1}{2}(1 - \gamma/\beta_1 - \gamma/\beta_2)$ and (x_1, x_2) is not a saddle point elsewhere.

Using the result of Theorem 2.1, as long as $\gamma > \beta_1\beta_2/(\sqrt{\beta_1} + \sqrt{\beta_2})^2$, Equation (2.3) would have no non-trivial solution and thus the first condition is satisfied. Note that, in some situations, a very high recovery rate of γ may not be possible. For example, we may not be able to implement an anti-virus software such that $\gamma > \beta_1\beta_2/(\sqrt{\beta_1} + \sqrt{\beta_2})^2$ with a limited budget or in limited time. In this case, we are still able to control the spreading of the virus if we can find a smaller and practical γ such that the second or the third condition is satisfied.

2.2 Generalized multi-state SIS model for complete graphs

Let us first analyse the general SIS model for $k \geqslant 2$ and that the network is a complete graph. Extension for the general network will be presented in later sections. For each $s \in \{0, 1, \ldots, k - 1\}$, let $x_s(t)$ be the fraction of nodes in state s at time t. Let $(x_0, x_1, \ldots, x_{k-1})$ be an equilibrium for the model. We can express the following system of differential equations:

$$\frac{\mathrm{d}x_{k-1}}{\mathrm{d}t} = \beta_{k-1}x_{k-2}x_{k-1} - \gamma x_{k-1}, \tag{2.5}$$

$$\frac{\mathrm{d}x_s}{\mathrm{d}t} = -\beta_{s+1}x_s x_{k-1} + \beta_s x_{s-1}x_{k-1} - \gamma x_s \quad \forall s \in \{1, \ldots, k - 2\}. \tag{2.6}$$

By equating $\mathrm{d}x_{k-1}/\mathrm{d}t = 0$, we have ($x_{k-1} = 0$) or ($x_{k-1} \neq 0$ and $x_{k-2} = \gamma/\beta_{k-1}$). We set $\mathrm{d}x_s/\mathrm{d}t = 0$ for any $1 \leqslant s \leqslant k - 2$, or $-\beta_{s+1}x_{k-1}x_s + \beta_s x_{k-1}x_{s-1} - \gamma x_s = 0$. This implies that we have the following relationship:

$$x_{s-1} = \left(\frac{\beta_{s+1}x_{k-1} + \gamma}{\beta_s x_{k-1}}\right) x_s = \left(\prod_{j=s}^{k-2} \frac{\beta_{j+1}x_{k-1} + \gamma}{\beta_j x_{k-1}}\right) x_{k-2}.$$

If $x_{k-1} \neq 0$, the condition $\sum_{i=0}^{k-1} x_i = 1$ is equivalent to

$$1 = x_{k-1} + \frac{\gamma}{\beta_{k-1}} + \sum_{s=1}^{k-2} \left(\prod_{j=s}^{k-2} \frac{\beta_{j+1}x_{k-1} + \gamma}{\beta_j x_{k-1}}\right) \frac{\gamma}{\beta_{k-1}},$$

which is a $(k - 1)$-dimension equation of x_{k-1}. By multiplying $\beta_1\beta_2 \cdots \beta_{k-1}x_{k-1}^{k-2}$ on both sides, we obtain

$$\beta_1\beta_2 \ldots \beta_{k-1}x_{k-1}^{k-2} = (\beta_1 x_{k-1} + \gamma) \cdots (\beta_{k-1}x_{k-1} + \gamma). \tag{2.7}$$

This relationship holds for any $k \geqslant 2$ (this can be formally shown via mathematical induction on k).

Let us now consider a special case where the infection rates are increasing geometrically with a growth rate $\alpha > 1$. In other words, $\beta_{k-1} = \alpha\beta_{k-2} = \cdots = \alpha^{k-2}\beta_1$. For positive β_1, Equation (2.7) is equivalent to $1/x_{k-1} = (1 + \gamma/\beta_1 x_{k-1}) \cdots (1 + \gamma/\beta_{k-1}x_{k-1})$. For simplicity, we make a homogeneous assumption where $\beta = \beta_1$. Substituting y with $1/\beta x_{k-1}$, we have $\beta y = (1 + \gamma y)(1 + (\gamma/\alpha)y) \cdots (1 + (\gamma/\alpha^{k-2})y)$.

Define $g_1(y) = \beta y$ and $g_2(y) = (1 + \gamma y)(1 + (\gamma/\alpha)y) \cdots (1 + (\gamma/\alpha^{k-2})y)$. These two functions of y are positive, monotone increasing and convex for $y > 0$. Since $g_2(y) = 0$ has only negative solutions $y = -\gamma, -\gamma\alpha, \ldots, -\gamma\alpha^{k-2}$. Thus, $g_1(y) = g_2(y)$ has at most two solutions. Moreover, for a fixed α, there is a *tipping point* β_t (or equivalently epidemic threshold). The equation has no solution if $\beta < \beta_t$ and has two solutions if $\beta > \beta_t$. Let us first present the *tipping point* analysis for a special case where $\alpha = 1$.

THEOREM 2.3 For the case where $\alpha = 1$ (i.e. the infection rates are homogeneous), $\beta \geqslant \beta_t = \gamma((k-1)^{k-1}/(k-2)^{k-2})$ is the necessary condition for the dynamic system (2.5) and (2.6) to have a non-trivial equilibrium. Or equivalently, $\beta < \beta_t$ is a sufficient condition for the fraction of nodes at the infected state to eventually converge to zero.

Proof. For the case where $\alpha = 1$, $g_1(y) = g_2(y)$ is equivalent to $\beta y = (1 + \gamma y)^{k-1}$. At the tipping point, i.e. $\beta = \beta_t$, $g_1(y) = g_2(y)$ has a unique solution. Hence, the slopes of $g_1(y)$ and $g_2(y)$ are the same when $g_1(y) = g_2(y)$, since $(\mathrm{d}/\mathrm{d}y)\beta_t y = \beta_t$ and $(\mathrm{d}/\mathrm{d}y)(1 + \gamma y)^{k-1} = \gamma(k-1)(1 + \gamma y)^{k-2}$ intersect at the point $y = (1/\gamma)[(\beta_t/\gamma(k-1))^{1/(k-2)} - 1]$. Substituting y with $(1/\gamma)[(\beta_t/\gamma(k-1))^{1/(k-2)} - 1]$, we obtain $(\beta_t/\gamma)[(\beta_t/\gamma(k-1))^{1/(k-2)} - 1] = [\beta_t/\gamma(k-1)]^{(k-1)/(k-2)}$. When β_t is positive, we can derive the expression of β_t as

$$\frac{\beta_t}{\gamma}\left[\left(\frac{\beta_t}{\gamma(k-1)}\right)^{1/(k-2)} - 1\right] = \left(\frac{\beta_t}{\gamma(k-1)}\right)^{(k-1)/(k-2)},$$

$$\frac{\beta_t}{\gamma}\left[\frac{\beta_t^{1/(k-2)}}{\gamma^{1/(k-2)}(k-1)^{1/(k-2)}} - 1\right] = \frac{\beta_t}{\gamma}\left[\frac{\beta_t^{1/(k-2)}}{r^{1/(k-2)}(k-1)^{(k-1)/(k-2)}}\right],$$

$$\beta_t^{1/(k-2)}\left[\frac{1}{\gamma^{1/(k-2)}(k-1)^{1/(k-2)}} - \frac{1}{\gamma^{1/(k-2)}(k-1)^{(k-1)/(k-2)}}\right] = 1,$$

$$\beta_t = \gamma\frac{(k-1)^{k-1}}{(k-2)^{k-2}}.$$

Thus, $\beta \geqslant \beta_t$ guarantees that $g_1(y) = g_2(y)$ will have at least one solution. Equivalently, Equation (2.7) will have at least one solution. Therefore, it is a necessary condition for the dynamic system (2.5) and (2.6) to have a non-trivial equilibrium.   □

It is important to node that, for $\alpha = 1$, $\beta \geqslant \beta_t$ is not sufficient for the dynamic system (2.5) and (2.6) to have a non-trivial equilibrium, since the solution x_{k-1} for Equation (2.7) may not be feasible (e.g. $x_{k-1} < 0$ or $x_{k-1} > 1$) and a feasible solution may not be a reachable equilibrium.

APPLICATION 2 (Vaccination strategy or product promotion strategy) Let us consider a computer virus outbreak in a network G. One can devise an effective vaccination strategy using Equations (2.5) and (2.6). For example, one can provide an anti-virus software with a detection rate of γ. If $\gamma > \beta((k-2)^{k-2}/(k-1)^{k-1})$, then the fraction of nodes in the infected state converges to zero for sufficiently large t. Furthermore, when k increases (i.e. the viruses need more phases to infect and activate a node), the threshold for γ decreases. Hence, it is easier to control the outbreak when we have a large k. On the other hand, if we want to promote a product in an online social network, then decreasing k makes our product easier to influence more customers.

Let us now consider the general case in which the infection rates are not homogeneous. For a given recovery rate γ and initial fraction of nodes in different states, it is straightforward to see that if the infection rates increase, the chance for the dynamic system (2.5) and (2.6) to reach a non-trivial equilibrium also increases. For example, suppose the infection rates for an epidemic increase, the chance for the epidemic outbreak also increases. Hence, we have the following corollary.

COROLLARY 2.1 Let $\beta_{\max} = \max_{i=1}^{k-1} \beta_i$, $\beta_{\max} \geqslant \gamma((k-1)^{k-1}/(k-2)^{k-2})$ be the necessary condition for the dynamic system (2.5) and (2.6) with heterogeneous infection rates to have a non-trivial solution.

2.3 Modelling generalized multi-state SIS model in general graphs

Now consider a general graph $G = (V, E)$ with a generalized SIS model of $k \geqslant 2$. Again, let A be the adjacency matrix of G. We make the following assumptions about the underlying graph G. If G is an undirected graph, we assume it is connected. If G is a directed graph, we assume it is weakly connected. (Since otherwise, we can analysis each weakly connected components separately.) Moreover, we assume that the graph contains more than one node, i.e. $|V| > 1$. For each $s \in \{0, 1, \ldots, k-1\}$, let $\langle x_s^{(i)}(t) \rangle$ be the average probability that node i is in state s at time t. Let $\langle x_{s_1}^{(i)}(t), x_{s_2}^{(j)}(t) \rangle$ be the average probability that node i is in state s_1 and node j is in state s_2. For a general graph G, the transmission rate is the rate at which infection will be transmitted between an infected individual and a susceptible individual. We denote the transmission rate by β_i' for each $i \in \{0, 1, \ldots, k-1\}$. Via the mean-field analysis, we obtain the following differential equation which describes the system dynamics:

$$\frac{d\langle x_{k-1}^{(i)} \rangle}{dt} = \beta_{k-1}' \sum_j A_{ji} \langle x_{k-2}^{(i)}, x_{k-1}^{(j)} \rangle - \gamma \langle x_{k-1}^{(i)} \rangle,$$

$$\frac{d\langle x_s^{(i)} \rangle}{dt} = -\beta_{s+1}' \sum_j A_{ji} \langle x_s^{(i)}, x_{k-1}^{(j)} \rangle + \beta_s' \sum_j A_{ji} \langle x_{s-1}^{(i)}, x_{k-1}^{(j)} \rangle - \gamma \langle x_s^{(i)} \rangle \quad \forall s \in \{1, \ldots, k-2\}.$$

Again, we have $x_0^{(i)} = 1 - \sum_{s=1}^{k-1} x_s^{(i)}$ for each node i.

Based on the *independence assumption*, we now provide an approximation to the above equation: $\langle x_{s_1}^{(i)}(t), x_{s_2}^{(j)}(t) \rangle = \langle x_{s_1}^{(i)}(t) \rangle \langle x_{s_2}^{(j)}(t) \rangle$. For simplicity of presentation, we omit the angle brackets. Then, for each node $i \in V$, the approximation equations are

$$\frac{dx_{k-1}^{(i)}}{dt} = \beta_{k-1}' x_{k-2}^{(i)} \sum_j A_{ji} x_{k-1}^{(j)} - \gamma x_{k-1}^{(i)}, \tag{2.8}$$

$$\frac{dx_s^{(i)}}{dt} = (\beta_s' x_{s-1}^{(i)} - \beta_{s+1}' x_s^{(i)}) \sum_j A_{ji} x_{k-1}^{(j)} - \gamma x_s^{(i)} \quad \forall s \in \{1, \ldots, k-2\}. \tag{2.9}$$

Note that the *independence assumption* we made here has been widely used in modelling and analyzing the epidemic threshold in general networks [3,7,8]. Moreover, in Section 3, we will illustrate via experiments that the numerical results given by solving the approximation equations agree very well with the simulation results.

For any given graph G and initial probability that node i is in state s, i.e. $x_s^{(i)}(0)$, we can use the above Equations (2.8) and (2.9) to numerically calculate the probability of each node in each state as a function of time.

2.4 Analysing the generalized SIS model for general undirected graphs

For general graphs, we are interested in finding the condition for the dynamic system (2.8) and (2.9) to reach a non-trivial equilibrium, i.e. $\lim_{t \to \infty} x_{k-1}^{(i)}(t) > 1$ for some $i \in V$. For simplicity, we let $x_s^{(i)} = \lim_{t \to \infty} x_s^{(i)}(t)$ to denote the probability of node i being in state s at equilibrium.

We first focus on a special case of our model where the infection rates are homogeneous, i.e.
$\beta'_1 = \beta'_2 = \cdots = \beta'_{k-1}$.

Prakash *et al.* [3] proved that for a series of virus propagation models operating on an arbitrary undirected graph with an adjacency matrix A and the largest eigenvalue λ_1, the virus will eventually be wiped out if $s < 1$ where s (*the effective strength*) is

$$s = \lambda_1 \cdot C_{\text{VPM}}$$

and C_{VPM} is an explicit constant dependent on the virus propagation model. Motivated by their work, we define the effective strength of our generalized SIS model for undirected graphs as follows.

DEFINITION 2.1 The effective strength of our generalized SIS model for undirected graphs is

$$s = \lambda_1 \frac{\beta'}{\gamma} \frac{(k-2)^{k-2}}{(k-1)^{k-1}}. \tag{2.10}$$

The following theorem gives the necessary condition for the existence of non-trivial equilibrium for the dynamic system (2.8) and (2.9).

THEOREM 2.4 For the generalized SIS model with homogeneous infection rates operating on an undirected connected graph with adjacency matrix A, the necessary condition for the dynamic system (2.8) and (2.9) to reach a non-trivial equilibrium is

$$s = \lambda_1 \frac{\beta'}{\gamma} \frac{(k-2)^{k-2}}{(k-1)^{k-1}} \geqslant 1, \tag{2.11}$$

where λ_1 is the largest eigenvalue of A.

Note that the generalized SIS model does not satisfy the general initial assumptions of the cascade model in [3] since the transitions between susceptible states depend on neighbours. Hence, the proof in [3] cannot be applied directly to prove Theorem 2.4. Before we present the proof of Theorem 2.4, let us first present some needed lemmas for the non-trivial equilibrium.

LEMMA 2.1 At equilibrium, we have

$$\beta' x^{(i)}_{k-2} \sum_j A_{ji} x^{(j)}_{k-1} = \gamma x^{(i)}_{k-1}, \tag{2.12}$$

$$\beta' (x^{(i)}_{s-1} - x^{(i)}_s) \sum_j A_{ji} x^{(j)}_{k-1} = \gamma x^{(i)}_s \quad \forall s \in \{1, \ldots, k-2\}. \tag{2.13}$$

Proof. The proof of this lemma can be shown by setting $dx^{(i)}_{k-1}/dt = 0$ and $dx^{(i)}_s/dt = 0$ $\forall s \in \{1, \ldots, k-2\}$. □

LEMMA 2.2 At the non-trivial equilibrium, we have $x^{(i)}_{k-1} > 0$ for all nodes i.

Proof. This can be proved by contradiction. Suppose that there exists a non-empty set of nodes $I \subset V$ such that $x_{k-1}^{(i)} = 0$ for all $i \in I$. For all nodes in $J = V - I$, $x_{k-1}^{(j)} > 0$. Based on the assumption that the graph is connected, there must exist a node $i \in I$ and a node $j \in J$ such that $A_{ji} > 0$.

Let t be the largest integer such that $x_t^{(i)} > 0$. Since $x_{k-1}^{(i)} = 0$, $t \leqslant k - 2$ holds. According to Equation (2.13), we have

$$\beta'(x_t^{(i)} - x_{t+1}^{(i)}) \sum_j A_{ji} x_{k-1}^{(j)} = \gamma x_{t+1}^{(i)}. \tag{2.14}$$

However, we have $\gamma x_{t+1}^{(i)} = 0$ and $\beta'(x_t^{(i)} - x_{t+1}^{(i)}) \sum_j A_{ji} x_{k-1}^{(j)} = \beta' x_t^{(i)} \sum_j A_{ji} x_{k-1}^{(j)} > 0$, which contradicts with Equation (2.14). The inequality holds since all nodes have positive degree in a connected undirected graph with more than one node. □

LEMMA 2.3 At the non-trivial equilibrium, for all i, we have

$$\frac{(\sum_j A_{ji} x_{k-1}^{(j)})^{k-1}}{x_{k-1}^{(i)}} = \left(\sum_j A_{ji} x_{k-1}^{(j)} + \frac{\gamma}{\beta'} \right)^{k-1}. \tag{2.15}$$

Proof. Lemma 2.2 implies $\sum_j A_{ji} x_{k-1}^{(j)} > 0$ holds for all i. Hence, Equation (2.12) can be written as

$$x_{k-2}^{(i)} = \frac{\gamma x_{k-1}^{(i)}}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}}.$$

According to Equation (2.13), for $1 \leqslant s \leqslant k - 2$, we have

$$x_{s-1}^{(i)} = \left(\frac{\gamma}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + 1 \right) x_s^{(i)} = \left(\frac{\gamma}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + 1 \right)^{k-s-1} x_{k-2}^{(i)}$$

$$= \left(\frac{\gamma}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + 1 \right)^{k-s-1} \frac{\gamma x_{k-1}^{(i)}}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}}.$$

Now, $\sum_{s=0}^{k-1} x_s^{(i)} = 1$ is equivalent to

$$1 = x_{k-1}^{(i)} + \frac{\gamma x_{k-1}^{(i)}}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + \sum_{s=1}^{k-2} \left(\frac{\gamma}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + 1 \right)^{k-s-1} \frac{\gamma x_{k-1}^{(i)}}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}},$$

$$\Leftrightarrow \frac{\sum_j A_{ji} x_{k-1}^{(j)}}{x_{k-1}^{(i)}} = \sum_j A_{ji} x_{k-1}^{(j)} + \frac{\gamma}{\beta'} \sum_{s=0}^{k-2} \left(\frac{\gamma}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + 1 \right)^s,$$

$$\Leftrightarrow \frac{1}{x_{k-1}^{(i)}} = \left(\frac{\gamma}{\beta' \sum_j A_{ji} x_{k-1}^{(j)}} + 1 \right)^{(k-1)}.$$

Multiplying $(\sum_j A_{ji} x_{k-1}^{(j)})^{k-1}$ on both sides, we obtain

$$\frac{(\sum_j A_{ji} x_{k-1}^{(j)})^{k-1}}{x_{k-1}^{(i)}} = \left(\sum_j A_{ji} x_{k-1}^{(j)} + \frac{\gamma}{\beta'} \right)^{k-1}. \tag{2.16}$$

□

We are now in the position to prove Theorem 2.4.

Proof of Theorem 2.4.   Since A^{T} is a symmetric real matrix, it has an orthonormal basis of eigenvectors, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{|V|}\}$ with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{|V|}$. Since we assume that the graph is connected and we know λ_1 is no less than the average degree of all nodes, we have $\lambda_1 > 0$. Suppose that the dynamic system (2.8–2.9) reaches a non-trivial equilibrium. Let $\mathbf{x}_{k-1} = (x_{k-1}^{(1)}, x_{k-1}^{(2)}, \ldots, x_{k-1}^{(|V|)})$. Then it can be written as a linear combination of eigenvectors of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{|V|}\}$.

$$\mathbf{x}_{k-1} = \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}. \tag{2.17}$$

Without loss of generality, we can assume that $c_{\ell} \geqslant 0$, since if can reverse the sign of $\mathbf{v}_{\ell}$ and it is still an eigenvector of A^T corresponding to eigenvalue λ_{ℓ}. Now, we have $A^T \mathbf{x}_{k-1} = \sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}$ and $\sum_j A_{ji} x_{k-1}^{(j)} = \sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}^{(i)}$. From Lemma 2.3, it follows that

$$\frac{(\sum_j A_{j1} x_{k-1}^{(j)})^{k-1}}{(\sum_j A_{j1} x_{k-1}^{(j)} + \gamma/\beta')^{k-1}} = x_{k-1}^{(1)}. \tag{2.18}$$

Let $\mathbf{v}_{\ell}^{(1)}$ be the first element of $\mathbf{v}_{\ell}$. Substituting $\sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}^{(1)}$ with $\sum_j A_{ji} x_{k-1}^{(j)}$ and substituting $\sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)}$ with $x_{k-1}^{(1)}$, Equation (2.18) becomes

$$\frac{(\sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}^{(1)})^{k-1}}{(\sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}^{(1)} + \gamma/\beta')^{k-1}} = \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)}. \tag{2.19}$$

Since $\lambda_1 \geqslant \lambda_{\ell}$ and $c_{\ell} \geqslant 0$ holds for all ℓ, we have $\lambda_1 \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)} \geqslant \sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}^{(1)}$. Applying Lemma 2.2, $x_{k-1}^{(1)} = \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)} > 0$ holds. Moreover, note that $h(x) = x^{k-1}/(x+t)^{k-1}$ is an increasing function of x for $x > 0$. Hence, from Equation (2.19), we can conclude $(\lambda_1 \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)})^{k-1}/(\lambda_1 \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)} + \gamma/\beta')^{k-1} \geqslant (\sum_{\ell} c_{\ell} \lambda_{\ell} \mathbf{v}_{\ell}^{(1)})^{k-1}/(\sum_{\ell} c_{\ell} \lambda_i \mathbf{v}_{\ell}^{(1)} + \gamma/\beta')^{k-1} = \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)}$.   It   follows $(1 + \gamma/\beta' \lambda_1 \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)})^{k-1} \leqslant 1/\sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)}$. Substituting y with $1/\beta' \lambda_1 \sum_{\ell} c_{\ell} \mathbf{v}_{\ell}^{(1)}$, it becomes

$$(1 + \gamma y)^{k-1} \leqslant \beta' \lambda_1 y. \tag{2.20}$$

If the dynamic system reaches a non-trivial equilibrium, there must exist a y such that the above inequality holds. It is easy to verify that $(1 + \gamma y)^{k-1}$ is a convex function of y. If there exists a y such that inequality (2.20) holds, then $(1 + \gamma y)^{k-1} = \beta' \lambda_1 y$ must have at least one solution.

In the proof of Theorem 2.3, we show that $\beta \geqslant \gamma((k-1)^{k-1}/(k-2)^{k-2})$ is a necessary condition for $(1 + \gamma y)^{k-1} = \beta y$ to have at least one solution. If follows naturally that $\lambda_1 \beta' \geqslant$

$\gamma((k-1)^{k-1}/(k-2)^{k-2})$ is a necessary condition for $(1+\gamma y)^{k-1} = \beta'\lambda_1 y$ to have at least one solution. Hence, the existence of y for Ineq. (2.20) implies $\lambda_1\beta' \geqslant \gamma((k-1)^{k-1}/(k-2)^{k-2})$, or equivalently, $s = \lambda_1(\beta'/\gamma)((k-2)^{k-2}/(k-1)^{k-1}) \geqslant 1$, which completes our proof. □

For the general case in which the infection rates are inhomogeneous, we have the following corollary.

COROLLARY 2.2 Let $\beta'_{\max} = \max_{i=1}^{k-1} \beta'_i$. For the generalized SIS model operating on an undirected connected graph with adjacency matrix A and largest eigenvalue λ_1, $\lambda_1(\beta'_{\max}/\gamma)((k-2)^{k-2}/(k-1)^{k-1}) \geqslant 1$ is a necessary condition for the dynamic system (2.8) and (2.9) to reach a non-trivial equilibrium.

2.5 *Analysing the generalized SIS model for general directed graphs*

Let us move on to find conditions for the dynamic system (2.8) and (2.9) to reach a non-trivial equilibrium for general directed graphs. We will first focus on the special case where the infection rates are homogeneous and all nodes in the directed graph have the same incident degrees. Then we present general results for general cases.

While analysing generalized multi-state SIS model for complete graphs, we use the mean-field approximation (2.8) and (2.9) and implicitly assumed that all nodes have the same probability of being in any given state s for $t = 0$. For consistency, we also make this assumption here. More precisely, we assume $x_s^{(1)}(0) = x_s^{(2)}(0) = \cdots = x_s^{(|V|)}(0)$ for any given state s and we say the dynamic system has homogeneous initial conditions for all nodes. We now provide another necessary condition for the dynamic system (2.8) and (2.9) to have a non-trivial solution.

THEOREM 2.5 For a directed graph $G = (V, E)$ with all nodes having the same incident degree d^-, the necessary condition for the dynamic system (2.8) and (2.9) with homogeneous infection rates β' and homogeneous initial conditions for all nodes to have a non-trivial equilibrium is $d^-\beta' \geqslant \gamma((k-1)^{k-1}/(k-2)^{k-2})$.

Proof. Since all nodes have same initial conditions and same incident degree, it is straightforward to see from Equations (2.8) and (2.9) that $x_s^{(1)}(t) = x_s^{(2)}(t) = \cdots = x_s^{(|V|)}(t)$ holds for any given $s \in \{1, 2, \ldots, k-1\}$ and time $t > 0$. Let $x_s(t) = x_s^{(1)}(t)$ for all $s \in \{1, 2, \ldots, k-1\}$. Then we have $\sum_j A_{ji} x_{k-1}^{(j)}(t) = d^- x_{k-1}(t)$ holds for all nodes i and any given time $t \geqslant 0$. Hence, finding results for Equations (2.8) and (2.9) is equivalent to finding results for the following Equations (2.21) and (2.22) and letting $x_s^{(1)}(t) = x_s^{(2)}(t) = \cdots = x_s^{(|V|)}(t) = x_s(t)$ for all s and $t \geqslant 0$.

$$\frac{dx_{k-1}}{dt} = d^-\beta' x_{k-2} x_{k-1} - \gamma x_{k-1}, \tag{2.21}$$

$$\frac{dx_s}{dt} = -d^-\beta' x_s + d^-\beta' x_{s-1} - \gamma x_s \quad \forall s \in \{1, \ldots, k-2\}. \tag{2.22}$$

Hence, the dynamic system (2.8) and (2.9) has a non-trivial solution if and only if the dynamic system (2.21) and (2.22) has non-trivial solution. Comparing Equations (2.21) and (2.22) with Equations (2.5) and (2.6), we can see that these two systems are equivalent if $\beta_1 = \beta_2 = \cdots = \beta_{k-1} = \beta = d^-\beta'$. Using the result from Theorem 2.3, we can conclude that $d^-\beta' \geqslant \gamma((k-1)^{k-1}/(k-2)^{k-2})$ is a necessary condition for the dynamic system (2.21) and (2.22) to have a non-trivial solution. This completes our proof. □

We have completed the derivation for the special case of our generalized SIS model with homogeneous infection rates operating on a graph with all nodes having the same incident degree. Let us now present results for the general case. For any directed graph G with adjacency matrix A, we can add some edges to G such that it becomes a graph G' with adjacency matrix A' satisfying $A'_{ij} = \max(A_{ij}, A_{ji})$ for all i, j. Moreover, suppose the maximum incident degree among all nodes in G is Δ^-; we can add some edges to G such that it becomes a graph $G'' = (V, E')$ with all nodes having the same incident degree Δ^-. Suppose that we are considering the spread of an epidemic. Since G' and G'' contains more edges ('possible contacts') than G, the epidemic is more likely to have an outbreak in G' and G'' than G. In other words, if the epidemic outbreaks in G at the end, it implies that the epidemic will also outbreak in G' and G''. Hence, it is straightforward to conclude that since G is a subgraph of G' and G'', for any given infection rates and recovery rates, the dynamic system (2.8) and (2.9) for G has a non-trivial solution which implies that the dynamic system (2.8) and (2.9) for G' and G'' also has a non-trivial solution.

Before showing the results for general directed graphs, let us first define the notion of effective strength for directed graphs.

DEFINITION 2.2 The effective strength of our generalized SIS model for directed graphs is

$$s = \min(\lambda_1(A'), \Delta^-)\frac{\beta'}{\gamma}\frac{(k-2)^{k-2}}{(k-1)^{k-1}}. \tag{2.23}$$

Now, for the discussion above, together with Theorems 2.4 and 2.5, we can establish the following corollary.

COROLLARY 2.3 For any directed graph $G = (V, E)$ with maximum incident degree Δ^- and adjacency matrix A, let $A'_{ij} = \max(A_{ij}, A_{ji})$ for all i, j, and $\lambda_1(A')$ be the largest eigenvalue of A'. For the generalized SIS model with homogeneous infection rates and homogeneous initial conditions for all nodes operating on G, the necessary condition for the existence of non-trivial equilibrium is

$$s = \min(\lambda_1(A'), \Delta^-)\frac{\beta'}{\gamma}\frac{(k-2)^{k-2}}{(k-1)^{k-1}} \geqslant 1. \tag{2.24}$$

For the general case where the infection rates are inhomogeneous, using the results from Corollaries 2.1 and 2.2, we have the following result.

COROLLARY 2.4 Let $\beta'_{\max} = \max_{i=1}^{k-1} \beta'_i$. For any directed graph $G = (V, E)$ with maximum incident degree Δ^- and adjacency matrix A, let $A'_{ij} = \max(A_{ij}, A_{ji})$ for all i, j, and $\lambda_1(A')$ be the largest eigenvalue of A'. For the generalized SIS model with homogeneous infection rates and homogeneous initial conditions for all nodes operating on G, the necessary condition for the existence of non-trivial equilibrium is

$$\min(\lambda_1(A'), \Delta^-)\beta'_{\max} \geqslant \gamma\frac{(k-1)^{k-1}}{(k-2)^{k-2}}. \tag{2.25}$$

3. Experimental results for the generalized SIS model

For the generalized SIS model, we conduct experiments based on both synthetic and real datasets and study the dynamics on fractions of states in a network. Note that our analysis of the condition for the existence of non-trivial solutions mainly focuses on models with homogeneous infection rates. In our experiments to verify these conditions, we also focus on models with homogeneous infection rates.
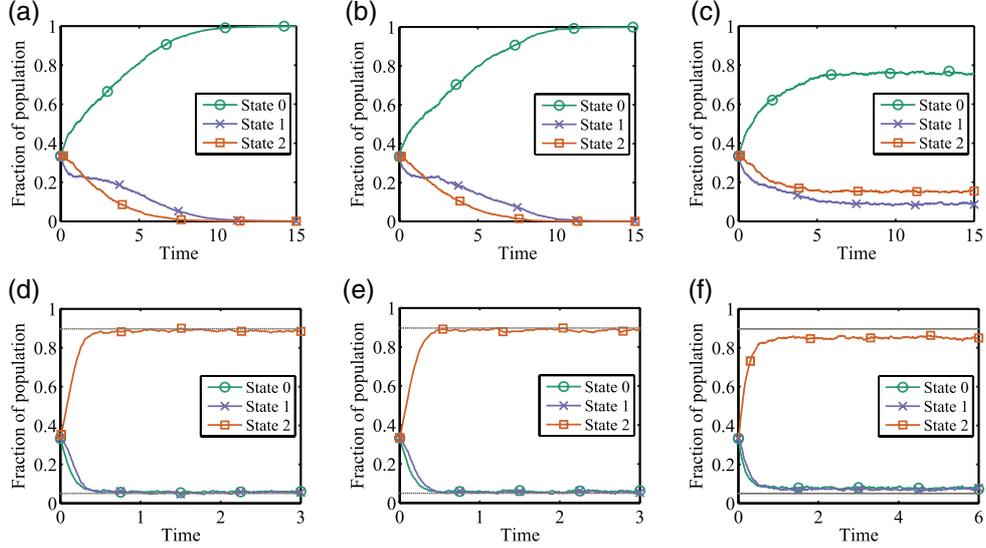
FIG. 3. Simulation results over synthetic datasets: The dynamics on fractions of nodes in different states over time for a complete graph, an Erdös–Rényi random graph and a power law random graph where $\gamma = 1$. (Initially all nodes have the same probability $1/k$ of being in any given state.) (a) K_{5000}, $\beta_1 = \beta_2 = 3$. (b) $G(5000, 0.05)$, $\beta'_1 = \beta'_2 = 3/250$. (c) $P(5000, 3, 250, 2500)$, $\beta'_1 = \beta'_2 = 3/250$. (d) K_{5000}, $\beta_1 = \beta_2 = 20$. (e) $G(5000, 0.05)$, $\beta'_1 = \beta'_2 = 20/250$. (f) $P(5000, 3, 250, 2500)$, $\beta'_1 = \beta'_2 = 20/250$.

3.1 Result over synthetic datasets

Our synthetic datasets include three undirected graphs: (i) a *complete graph* K_N with N nodes, (ii) an *Erdös–Rényi random graph* $G(N, p)$ with N nodes and each edge is included in the graph with probability p independent from every other edge and (iii) a *random power law graph* $P(N, \theta, d, m)$ with N nodes, the exponent θ, the average degree d and the maximum degree m [9].

Figure 3 shows the simulation results of the ternary model in a complete graph, an Erdös–Rényi random graph and a random power law graph. Note that although our analysis of ternary model focuses on the case (i), our results can be applied to the case (ii) with some constant factor with respect to $d = Np$, the expected average degree of the Erdös–Rényi random graph $G(N, p)$. Figure 3 compares the dynamics for the ternary model with different β_1 and β_2 values. Figure 3(a) shows that if there is no non-trivial stationary equilibrium, then (x_0, x_1, x_2) converges to $(1, 0, 0)$ for large t. But if there is a stationary equilibrium, then there is a possibility that (x_1, x_2) converges to another point. In Fig. 3(d), the dotted lines represent a stationary non-trivial equilibrium $(x_1, x_2) = (0.05, 0.897)$ and (x_1, x_2) converges to this non-trivial equilibrium. Figure 3(b) and (e) show that the analysis holds for $\beta'_1 = \beta_1/d$ and $\beta'_2 = \beta_2/d$ $(d = Np = 250)$. We like to point out that, for the case (iii), our method is successful in stating the equilibrium condition with $\beta'_1 = \beta_1/d$ and $\beta'_2 = \beta_2/d$ $(d = 250)$; this is shown both in Fig. 3(c) and (f).

Figure 4 shows the numerical results given by solving Equation (2.1–2.2) for the complete graph and Equations (2.8) and (2.9) for general graphs. Comparing Figs 3 and 4, we can see that the numerical results based on mean-field analysis agree very well with the simulation results. From now on, for the synthetic graphs under study, we use numerical results as the approximation of the simulation results.

Figures 5 and 6 illustrate our results about the necessary condition for the existence of non-trivial equilibrium. We use the same 'take-off' plots used by Prakash *et al.* [3] to illustrate our experimental
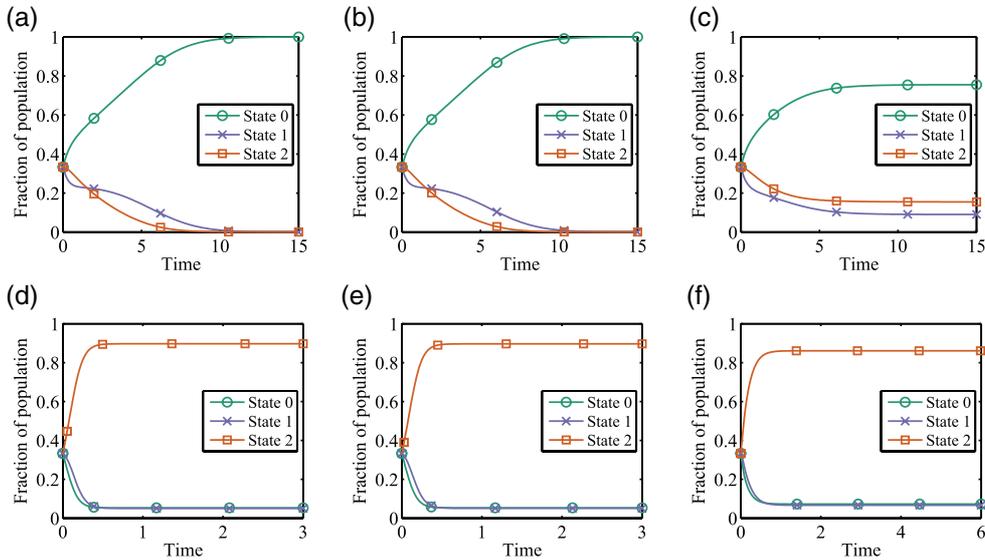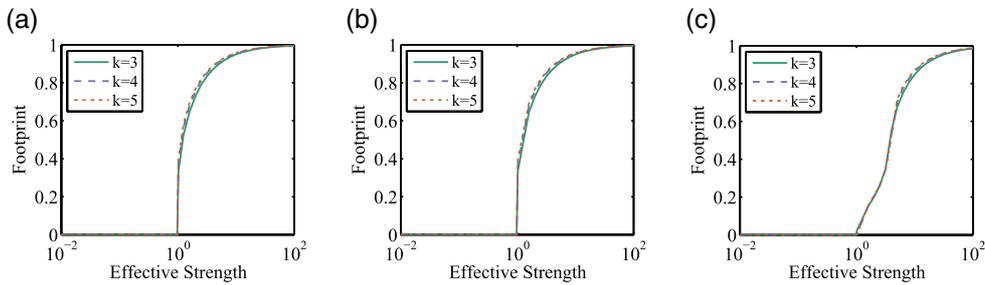
FIG. 4. Numerical results over synthetic datasets. The dynamics on fractions of nodes in different states over time for a complete graph, an Erdös–Rényi random graph and a random power law graph where $\gamma = 1$. (The initial probability of being in state s for each node i: $x_s^{(i)}(0) = 1/k$.) (a) K_{5000}, $\beta_1 = \beta_2 = 3$. (b) $G(5000, 0.05)$, $\beta_1' = \beta_2' = 3/250$. (c) $P(5000, 3, 250, 2500)$, $\beta_1' = \beta_2' = 3/250$. (d) K_{5000}, $\beta_1 = \beta_2 = 20$. (e) $G(5000, 0.05)$, $\beta_1' = \beta_2' = 20/250$. (f) $P(5000, 3, 250, 2500)$, $\beta_1' = \beta_2' = 20/250$.

FIG. 5. Numerical Results over synthetic datasets: 'Take-off' plots, Footprint vs. Effective strength (lin-log) for a complete graph, an Erdös–Rényi random graph and a power law random graph. The tipping point matches our necessary condition analysis in all cases. (The initial probability of being in state s for each node i: $x_s^{(i)}(0) = 1/k$.) (a) K_{5000}. (b) $G(5000, 0.05)$. (c) $P(5000, 3, 250, 2500)$.

results. The 'footprint' [3] is a measurement of the extent of infection. We define the footprint in our numerical results as the fraction of nodes being infected when the dynamic system reaches its equilibrium.

Theorem 2.4 suggests that if the effective strength $s < 1$, there exists no non-trivial equilibrium and the fraction of nodes being infected converges to zero in all models. In Figs 5 and 6, the system always reaches the trivial equilibrium when $s < 1$. In Fig. 5, the initial fraction of nodes being infected is $1/k$, the footprint of infection suddenly jumps at $s = 1$. This implies that, for the case where the initial fraction of nodes being infected is sufficiently large, $s = 1$ is a tipping point for the phase transition. If
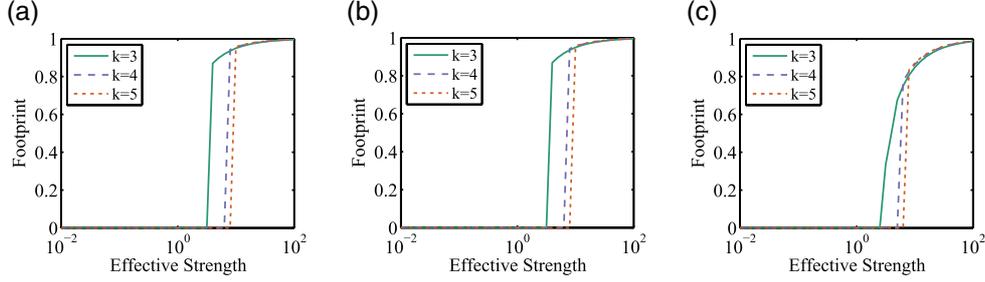
FIG. 6. Numerical Results over synthetic datasets: 'Take-off' plots, Footprint vs. Effective strength (lin-log) for a complete graph, an Erdös–Rényi random graph and a power law random graph. The tipping point matches our necessary condition analysis in all cases. (The initially probability of being in different states for each node i: $x_0^{(i)} = 0.99$, $x_{k-1}^{(i)} = 0.01$.) (a) K_{5000}. (b) $G(5000, 0.05)$. (c) $P(5000, 3, 250, 2500)$.

the effective strength is slightly above this threshold, the system will reach a non-trivial equilibrium. Under the threshold, there does not exist any non-trivial equilibrium. However, the footprint of infection does not suddenly jump at $s = 1$ as shown in Fig. 6. It coincides with our discussion in Section 2.4 that $s > 1$ is only a necessary condition for the system to reach a non-trivial equilibrium, not a sufficient one. For the system to reach a non-trivial equilibrium, the smaller the initial fraction of nodes being infected, the larger is the effective strength s needed. From Fig. 6, we can also see that, for the same set of initial conditions, the larger the value of k, the larger is the effective strength s needed for the existence of non-trivial equilibrium.

3.2 *Result over real datasets*

Our real datasets include three networks: (i) A *Facebook-like social network* originating from an online community for students at the University of California, Irvine [10]. The dataset contains $1,899$ users and $20,296$ directed edges. The largest weakly connected component contains $1,893$ users. (ii) An *Enron email network* that covers all the email communication from around half a million emails [11,12]. The dataset contains $36,692$ email addresses and $183,831$ undirected edges. The largest connected component contains $33,696$ email addresses. (iii) An *Epinions social network* of the who-trust-whom relationships from a consumer review site Epinions.com [13]. This network contains $508,837$ directed 'trust' relationships among $75,879$ users. All nodes are in the largest weakly connected component.

Figure 7 shows the simulation results and the numerical results based on the *Facebook-like social network*. The numerical results are given by solving Equations (2.8) and (2.9). Comparing Fig. 7(a, b) with Fig. 7(c, d), we can see that the numerical results based on mean-field analysis agree very well with the simulation results.

Figures 8 and 9 illustrate our results about the necessary condition for the existence of non-trivial equilibrium using the 'take-off' plots [3]. We run each simulation for 50 units of time and took the average of 20 runs. The 'footprint' of simulation results is defined as the maximum fraction of nodes infected during the second half of the simulation.

For the undirected *Enron email network*, applying Theorem 2.4, we know that the effective strength $s = \lambda_1(A)(\beta'/\gamma)((k-2)^{k-2}/(k-1)^{k-1}) \geqslant 1$ is a necessary condition for the dynamic system (2.8) and 2.9) to reach a non-trivial equilibrium. For directed networks, let $A'_{ij} = \max(A_{ij}, A_{ji})$ and Δ^- be the maximum incident degree. For the *Facebook-like social network*, we have $\lambda_1(A') = 48.1431$ and $\Delta^- = 137$. For the *Epinions social network*, we have $\lambda_1(A') = 106.5279$ and $\Delta^- = 3035$. Hence, we have
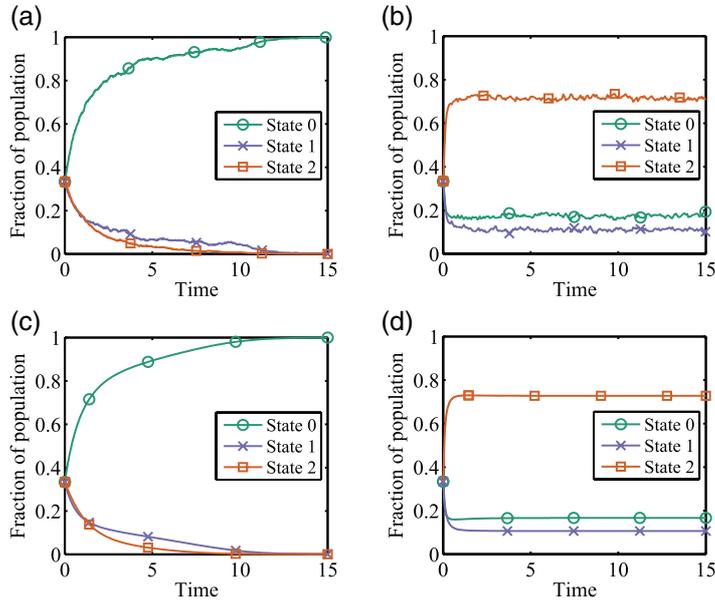
FIG. 7. Results over the *Facebook-like* network: Fraction of nodes being in different state over time where $k = 3$ and $\gamma = 1$. (a, b) Plot of simulation results. (c, d) Plot of numerical results. (Initially all nodes have the same probability $1/k$ of being in any given state.) (a) $\beta'_1 = \beta'_2 = 0.01$. (b) $\beta'_1 = \beta'_2 = 2$. (c) $\beta'_1 = \beta'_2 = 0.01$. (d) $\beta'_1 = \beta'_2 = 2$.
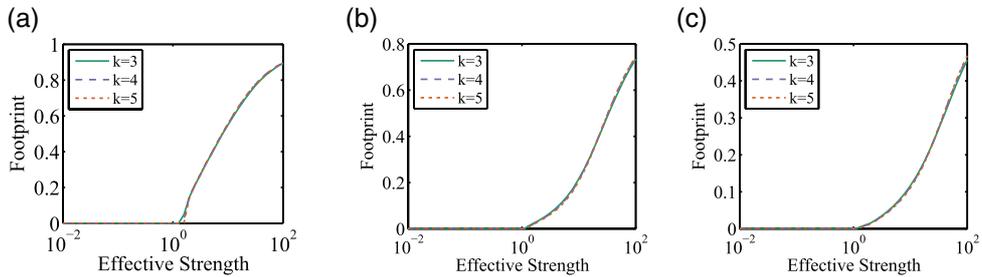

FIG. 8. Simulation results over three real networks: 'Take-off' plots, Footprint vs. Effective strength (lin-log) for the complete graph, Erdös–Rényi random graph and power law random graph. The tipping point matches our necessary condition analysis in all cases. (Initially all nodes have same probability $1/k$ of being in any given state.) (a) Facebook-like social network. (b) Enron email network. (c) Epinions social network.

$\lambda_1(A') < \Delta^-$ for both networks. Applying Corollary 2.3, one can conclude that the effective strength $s = \lambda_1(A')(\beta'/\gamma)((k-2)^{k-2}/(k-1)^{k-1}) \geqslant 1$ is a necessary condition for the dynamic system (2.8–2.9) to reach a non-trivial equilibrium.

The tipping point in Figs 8 and 9 matches our necessary condition analysis. For effective strength $s < 1$, the fraction of nodes being infected converges to zero for large t. For effective strength $s > 1$, the dynamic system may reach a non-trivial equilibrium. Moreover, whether the system will reach a non-trivial equilibrium or not depends not only on the effective strength, but also on the initial fraction of nodes being infected.
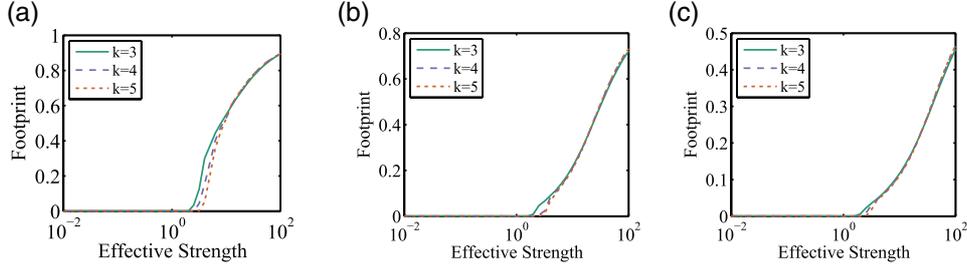
FIG. 9. Simulation results over three real networks: 'Take-off' plots, Footprint vs. Effective strength (lin-log) for the *Facebook-like social network*, the *Enron email network* and the*Epinions social network*. The tipping point matches our necessary condition analysis in all cases. (Initially all nodes have probability 0.01 of being in the infected state and probability 0.99 of being in state 0.) (a) Facebook-like social network. (b) Enron email network. (c) Epinions social network.

4. General SIS model with competing sources

We like to note that some of the previous work only considered one contagion source in a network. In practice, however, there might be more than one kind of contagions, ideas or behaviours spreading at the same time. Here, we consider two *competing sources*: *a* and *b*, which simultaneously spread their influence in the same network. We use the generalized SIS model discussed in Section 2 to model these two competing sources. In the following analysis and without loss of generality, we assume source *b* is dominant over source *a* such that even when a node is on its way to become activated by source *a*, it is still possible for this node to be influenced by its neighbours activated by source *b* and become activated by source *b* eventually. Note that the converse is not true, or source *a* has no such power over *b*. One application of such a model is to consider a spreading of virus *a* and one can eliminate the spreading of virus *a* by an antidote (model as source *b*). Let us first present the formal analysis of a ternary model in a large complete graph; then we formulate the model with two competing sources under a general graph.

4.1 *Ternary model in a large complete graph*

We consider a model with two competing sources '*a*' and '*b*' as depicted in Fig. 10. In this model, the underlying network is a large complete graph $G = (V, E)$. Any node $v \in V$ can be in one of the five states: $\{0, a_1, a_2, b_1, b_2\}$. Nodes in state a_2 and b_2 are in the activation state for a and b, respectively. Nodes in state a_2 (or b_2) can change the state value of its neighbours, say node s, which is in state 0 or a_1 (0 or b_1), to state a_1 or a_2 (b_1 or b_2). Furthermore, to represent the dominant behaviour of source b, nodes in state b_2 can change their neighbouring nodes in state a_1 to state b_1 with a non-zero probability. Nodes in state a_1 or a_2 can independently recover with a rate γ_a (recovery rate). Similarly, nodes in state b_1 or b_2 can recover with a rate γ_b. Assume that the underlying network is a complete graph. For each state $s \in \{0, a_1, a_2, b_1, b_2\}$, let $x_s(t)$ be the fraction of nodes in state s at time t. We can express the system dynamics using the following differential equations:

$$\frac{\mathrm{d}x_{a_1}}{\mathrm{d}t} = \alpha_1 x_0 x_{a_2} - \alpha_2 x_{a_1} x_{a_2} - \lambda x_{a_1} x_{b_2} - \gamma_a x_{a_1}, \tag{4.1}$$

$$\frac{\mathrm{d}x_{a_2}}{\mathrm{d}t} = \alpha_2 x_{a_1} x_{a_2} - \gamma_a x_{a_2}, \tag{4.2}$$
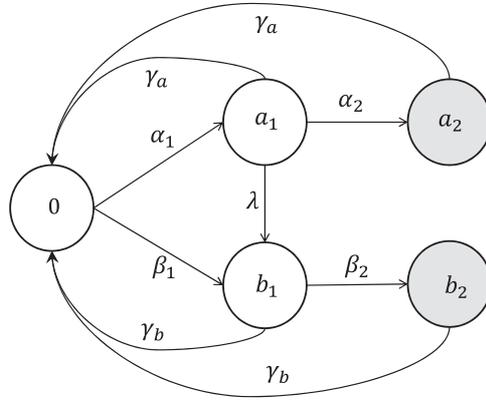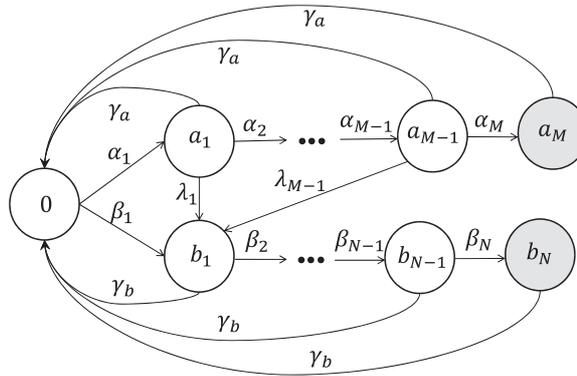
FIG. 10. The ternary SIS model with two competing sources.

FIG. 11. The generalized SIS model with two competing sources.

$$\frac{\mathrm{d}x_{b_1}}{\mathrm{d}t} = \beta_1 x_0 x_{b_2} - \beta_2 x_{b_1} x_{b_2} + \lambda x_{a_1} x_{b_2} - \gamma_b x_{b_1}, \tag{4.3}$$

$$\frac{\mathrm{d}x_{b_2}}{\mathrm{d}t} = \beta_2 x_{b_1} x_{b_2} - \gamma_b x_{b_2}, \tag{4.4}$$

$$x_0(t) = 1 - x_{a_1}(t) - x_{a_2}(t) - x_{b_1}(t) - x_{b_2}(t). \tag{4.5}$$

For the above system, there does not exist a closed-form solution, but one can solve it numerically so as to understand the dynamics of these two competing sources. In the following section, we will present our numerical results.

4.2 *General multi-state model in a large complete graph*

Now we consider a general SIS with two competing sources. Let $\mathcal{M} \geqslant 2$ and $\mathcal{N} \geqslant 2$ be the number of additional states for sources a and b. Since there is one additional initial state 0, the total number of

states is $\mathcal{M} + \mathcal{N} + 1$. For example, in the ternary model, we have $\mathcal{M} = \mathcal{N} = 2$. Figure 11 depicts the state-transition diagram of the general multi-state model.

Observe that when a node i is activated by the dominant source b and one of its neighbouring nodes j is in state a_i ($1 \leqslant i \leqslant \mathcal{M} - 1$), then the node i can change node j from state a_i to state b_1 with probability λ_i per unit time. In some real-world cases, for a node in state a_k ($1 \leqslant k \leqslant \mathcal{N} - 1$), a higher value of k implies that the node is closer to being activated by the source a. In this case, we may have $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{\mathcal{M}-1}$. For example, let us consider a consumer purchase decision process and suppose $\mathcal{M} = 3$. Then a_1 is the stage that a customer has heard about a new product and a_2 is the stage that he is planning to make a purchase. For this example, a customer in stage a_1 is easier to change his mind if he is exposed to a more dominant product, say b. In other words, we can assume that $\lambda_1 \geqslant \lambda_2$. Note that $x_0 = 1 - \sum_{i=1}^{\mathcal{M}} x_{a_i} - \sum_{j=1}^{\mathcal{N}} x_{b_j}$. The system dynamics can be specified by the following equations:

$$\frac{dx_{a_1}}{dt} = \alpha_1 x_0 x_{a_\mathcal{M}} - \alpha_2 x_{a_1} x_{a_\mathcal{M}} - \lambda_1 x_{a_1} x_{b_\mathcal{N}} - \gamma_a x_{a_1}, \tag{4.6}$$

$$\frac{dx_{a_i}}{dt} = \alpha_i x_{a_{i-1}} x_{a_\mathcal{M}} - \alpha_{i+1} x_{a_i} x_{a_\mathcal{M}} - \lambda_i x_{a_i} x_{b_\mathcal{N}} - \gamma_a x_{a_i} \quad \forall i \in \{2, \ldots, \mathcal{M} - 1\}, \tag{4.7}$$

$$\frac{dx_{a_\mathcal{M}}}{dt} = \alpha_\mathcal{M} x_{a_{\mathcal{M}-1}} x_{a_\mathcal{M}} - \gamma_a x_{a_\mathcal{M}}, \tag{4.8}$$

$$\frac{dx_{b_1}}{dt} = \beta_1 x_0 x_{b_\mathcal{N}} - \beta_2 x_{b_1} x_{b_\mathcal{N}} + \sum_{\ell=1}^{\mathcal{M}-1} \lambda_\ell x_{a_\ell} x_{b_\mathcal{N}} - \gamma_b x_{b_1}, \tag{4.9}$$

$$\frac{dx_{b_j}}{dt} = \beta_j x_{b_{j-1}} x_{b_\mathcal{N}} - \beta_{j+1} x_{b_j} x_{b_\mathcal{N}} - \gamma_b x_{b_j} \quad \forall j \in \{2, \ldots, \mathcal{N} - 1\}, \tag{4.10}$$

$$\frac{dx_{b_\mathcal{N}}}{dt} = \beta_\mathcal{N} x_{b_{\mathcal{N}-1}} x_{b_\mathcal{N}} - \gamma_b x_{b_\mathcal{N}}. \tag{4.11}$$

It is easy to see that the ternary model (where $\mathcal{M} = \mathcal{N} = 2$) discussed in Section 4.1 is in fact a special case of this general model, where $\mathcal{M} = \mathcal{N} = 2$.

4.3 *Multi-state model in a general graph*

When the underlying network is a complete graph, we assume that contact is possible with the entire population. However, for general graphs, only an activated node can influence its neighbours. In this model, the transmission rate is the rate at which a source is being transmitted between two nodes, one activated and one non-activated, and they are connected by an edge in the graph. In a complete graph, the transmission rate is the rate of contacts between an activated node and all others, whereas in a general graph it is the rate of contacts between neighbouring nodes. We denote the transmission rate here by α_i' ($1 \leqslant i \leqslant \mathcal{M}$), β_j' ($1 \leqslant j \leqslant \mathcal{N}$) and λ_ℓ' ($1 \leqslant \ell \leqslant \mathcal{M} - 1$). For example, assume that node i and node j are connected. Suppose node i is in state $b_\mathcal{N}$ (activated by source b) and node j is in state 0, node j can change from state 0 to b_1 with rate β_1'. Suppose another node k is in state a_1. The rate that node k will change to state b_1 is λ_1'. Noted that the transmission rates in a complete graph and a general graph are different from each other.

Let $G = (V, E)$ be the underlying general graph, where A is the adjacency matrix of G. Let $\mathcal{M} \geqslant 2$ and $\mathcal{N} \geqslant 2$ be the number of additional states of sources a and b, respectively. For each $s \in \{0, a_1, \ldots, a_\mathcal{M}, b_1, \ldots, b_\mathcal{N}\}$, let $x_s^{(i)}(t)$ be the average probability that node i is in state s at time t. Let

$\langle x_{s_1}^{(i)}(t), x_{s_2}^{(j)}(t)\rangle$ be the average probability that node i is in state s_1 and node j is in state s_2 at time t. We also use the approximation that $\langle x_{s_1}^{(i)}, x_{s_2}^{(j)}\rangle = x_{s_1}^{(i)} x_{s_2}^{(j)}$ to close the equations at the level of a one-variable average. We can express the system dynamics as follows:

$$\frac{dx_{a_1}^{(i)}}{dt} = \alpha_1' x_0^{(i)} \sum_j A_{ji} x_{a_\mathcal{M}}^{(i)} - \alpha_2' x_{a_1}^{(i)} \sum_j A_{ji} x_{a_\mathcal{M}}^{(i)} - \lambda_1' x_{a_1}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \gamma_a x_{a_1}^{(i)}, \tag{4.12}$$

$$\frac{dx_{a_k}^{(i)}}{dt} = \alpha_k' x_{a_{k-1}}^{(i)} \sum_j A_{ji} x_{a_\mathcal{M}}^{(i)} - \alpha_{k+1}' x_{a_k}^{(i)} \sum_j A_{ji} x_{a_\mathcal{M}}^{(i)} - \lambda_k' x_{a_k}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \gamma_a x_{a_k}^{(i)} \quad 2 \leqslant k \leqslant \mathcal{M} - 1, \tag{4.13}$$

$$\frac{dx_{a_\mathcal{M}}}{dt} = \alpha_\mathcal{M}' x_{a_{\mathcal{M}-1}}^{(i)} \sum_j A_{ji} x_{a_\mathcal{M}}^{(i)} - \gamma_a x_{a_\mathcal{M}}^{(i)}, \tag{4.14}$$

$$\frac{dx_{b_k}^{(i)}}{dt} = \beta_1' x_0^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \beta_2' x_{b_k}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} + \sum_{\ell=1}^{\mathcal{M}-1} \lambda_\ell' x_{a_\ell}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \gamma_b x_{b_k}^{(i)}, \tag{4.15}$$

$$\frac{dx_{b_k}}{dt} = \beta_k' x_{b_{k-1}}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \beta_{k+1}' x_{b_k}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \gamma_b x_{b_k}^{(i)} \quad 2 \leqslant k \leqslant \mathcal{N} - 1, \tag{4.16}$$

$$\frac{dx_{b_\mathcal{N}}^{(i)}}{dt} = \beta_\mathcal{N}' x_{b_{\mathcal{N}-1}}^{(i)} \sum_j A_{ji} x_{b_\mathcal{N}}^{(i)} - \gamma_b x_{b_\mathcal{N}}^{(i)}, \tag{4.17}$$

$$x_0^{(i)} = 1 - \sum_{i=1}^{\mathcal{M}} x_{a_i}^{(i)} - \sum_{j=1}^{\mathcal{N}} x_{b_j}^{(i)}. \tag{4.18}$$

The above derivation is equivalent to the dynamic system (4.6–4.11) for complete graphs when $\alpha_i = (|V| - 1)\alpha_i'$ for all $1 \leqslant i \leqslant \mathcal{M}$, $\lambda_\ell = (|V| - 1)\lambda_\ell'$ for all $1 \leqslant \ell \leqslant \mathcal{M} - 1$ and $\beta_j = (|V| - 1)\beta_j'$ for all $1 \leqslant j \leqslant \mathcal{N}$.

5. Experimental results for competing sources

In this section, we provide numerical results for the generalized SIS model with two competing sources. As in Section , we consider networks of (i) a complete graph K_N with N nodes, (ii) a *Erdös–Rényi random graph* $G(N, p)$ with N nodes and each edge is included in the graph with probability p independent from every other edge and (iii) a random power law graph $P(N, \theta, d, m)$. For the initial condition, the initial state value of each node is chosen uniformly and independently according to a given initial rate. *Results for ternary model*: The simulation results of the ternary model in a complete graph, an Erdös–Rényi random graph and a random power law graph are illustrated in Fig. 12. One can see that when transmission rates α_i and β_i are not sufficiently large, the epidemic will eventually die out. In Fig. 12(c–d), we have $\alpha = 250\alpha'$, $\beta = 250\beta'$ and $\lambda = 250\lambda'$, where 250 is the expected average degree of nodes in the two random graphs. Thus, the expected number of contacts of an activated node remains the same. We also observe a similar phase transition and equilibrium in all of these three networks.

Figure 13 depicts the numerical results given by solving Equations (4.1–4.5) for a complete graph and Equations (4.12)–(4.18) for a general graph. Comparing Fig. 13 with Fig. 12, one can observe that
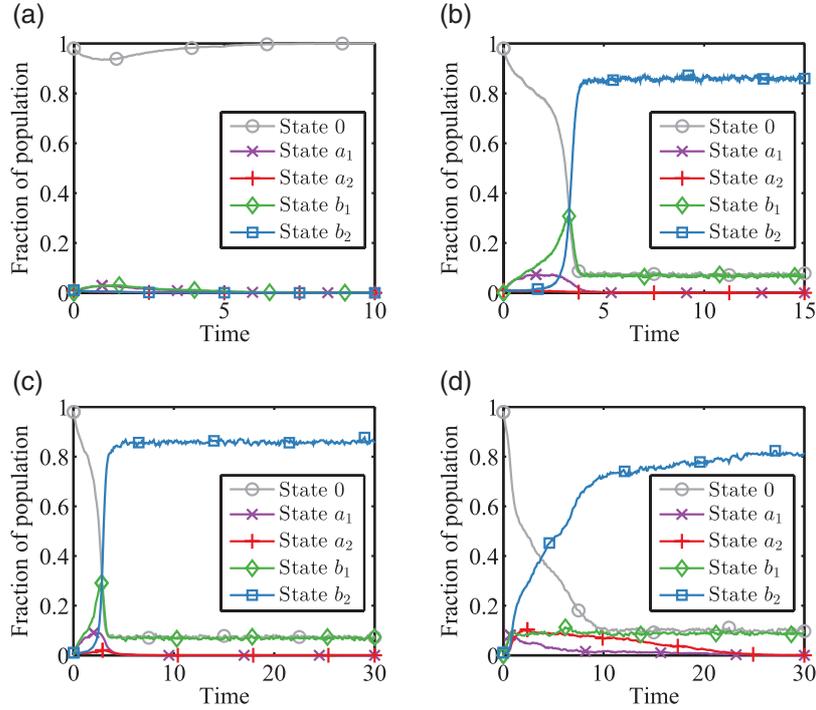
FIG. 12. Simulation results: The dynamics of x_0, x_{a_1}, x_{a_2}, x_{b_1} and x_{b_2} over time for a complete graph, an Erdös–Rényi random graph and a random power law graph where $\gamma_a = \gamma_b = 1$. (Initially, each node has probability 0.01 of being in state a_2, probability 0.01 of being in state b_2 and probability 0.98 of being in state 0.) (a) K_{5000}, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 8$, $\lambda = 2$. (b) K_{5000}, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 16$, $\lambda = 2$. (c) $G(5000, 0.05)$, $\alpha'_1 = \alpha'_2 = 16/250$, $\beta'_i = \beta'_2 = 16/250$, $\lambda' = 2/250$. (d) $P(5000, 3, 250, 2500)$, $\alpha'_1 = \alpha'_2 = 16/250$, $\beta'_i = \beta'_2 = 16/250$, $\lambda' = 2/250$.

the numerical results of solving the differential equations are very close to the numerical results given by simulation. As we discussed in Section 4.3, Equations (4.12–4.18) is an approximation of the dynamic system for general graphs. From the experimental results, we can see that the approximation is very close to the simulation results.

From now on, we show only the experiments based on complete graphs. For Erdös–Rényi random graphs and power law graphs, the results could also be applied. We use numerical results given by Equations (4.6–4.11). Note that, for complete graphs, the numerical results do not depend on the number of nodes. So we do not need to specify the number of nodes for the numerical results for the complete graph.

Impact of delay in deploying source b: Fig. 14 depicts the numerical results for which $x_{a_2}(0) = 2x_{b_2}(0) = 0.02$. By setting different initial fractions of infected nodes for sources a and b, we can examine the impact of delay in the phase-transition process. For example, if $x_{a_2}(0) > x_{b_2}(0)$, we can assume that source b is introduced later than a. In Fig. 14(a), we can see that, for product b, if $\beta_{1,2}$ and λ are not large enough, it cannot compete well with product a. Eventually, the fraction of nodes in state b_2 will approach zero. On the other hand, for a product b, if it is far more superior than product a (i.e., λ and $\beta_{1,2}$ are sufficiently large, or γ_b is small), potential buyers can be easily persuaded
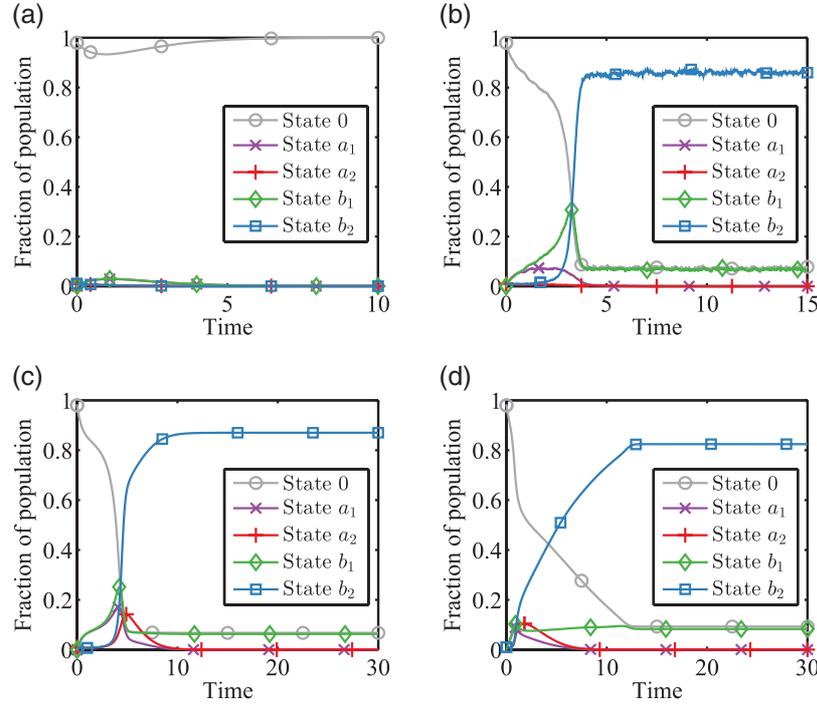
FIG. 13. Numerical results: The dynamics of x_0, x_{a_1}, x_{a_2}, x_{b_1} and x_{b_2} over time for a complete graph, an Erdös–Rényi random graph and a random power law graph where $\gamma_a = \gamma_b = 1$. (The initial probability of being in different states for each node i: $x_0^{(i)}(0) = 0.98$, $x_{a_2}^{(i)}(0) = 0.01$ and $x_{b_2}^{(i)}(0) = 0.01$.) (a) K_{5000}, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 8$, $\lambda = 2$. (b) K_{5000}, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 16$, $\lambda = 2$. (c) $G(5000, 0.05)$, $\alpha'_1 = \alpha'_2 = 16/250$, $\beta'_1 = \beta'_2 = 16/250$, $\lambda' = 2/250$. (d) $P(5000, 3, 250, 2500)$, $\alpha'_1 = \alpha'_2 = 16/250$, $\beta'_1 = \beta'_2 = 16/250$, $\lambda' = 2/250$.

to eventually adopt product b. Figure 14(b–d) correspond to these situations. If product b is superior, it will be the dominant source even if it is introduced to the network at a much later stage than product a.

Impact of λ: Fig. 15 shows that λ plays an important role for source b to be dominant. Suppose the fraction of nodes in state a_2 equals the fraction of nodes in state b_2 initially. Intuitively speaking, a larger λ means nodes in state a_1 have a higher probability of being changed to state b_1. From the figure, we can observe that the larger the λ is, the sooner the network reaches its equilibrium. In addition, suppose we fix $\alpha_{1,2}$ and $\beta_{1,2}$ such that at equilibrium the fraction of nodes in state a_2 is zero. In this scenario, the value of λ only has the impact on the time taken to reach the equilibrium, and it has a negligible impact on the final distribution of nodes in different states.

Impact of $\mathcal{M}$ *and* $\mathcal{N}$: For a multi-state model with one source, as we have discussed in Section 2, when k (i.e. the number of steps to activate a node from state 0) increases, the threshold for γ decreases so as to prevent a phase transition. In other words, if the recovery rate γ remains the same, and if we want to maximize the influence of a source, it is important for us to decrease k.

For a multi-state model with two competing sources, similar conclusions can be made. Since source b is the dominant one, we could assume that $\mathcal{N} \leqslant \mathcal{M}$ (i.e. it takes less phases for a node to be activated
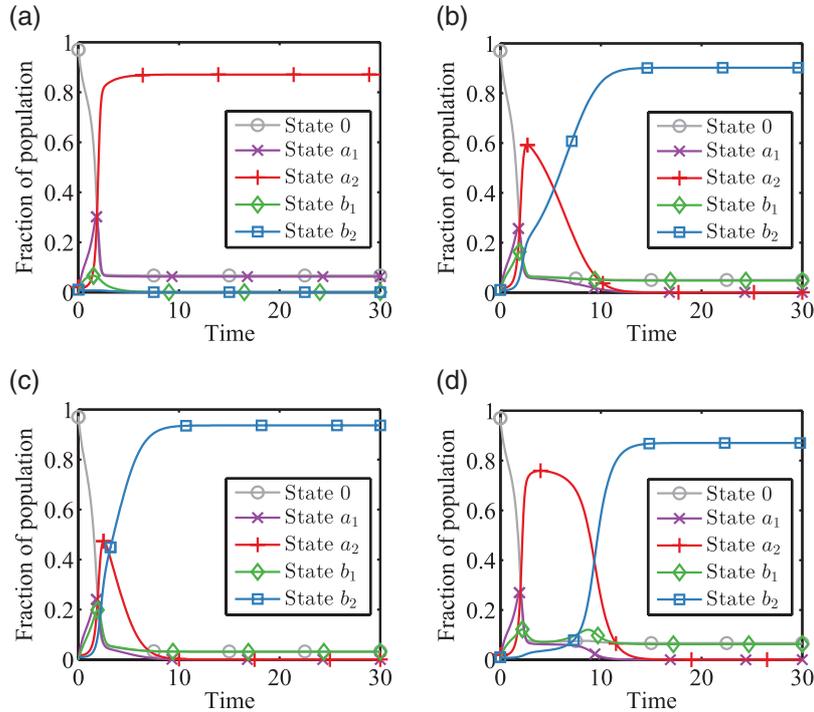
FIG. 14. Numerical results over a complete graph showing the impact of delay in deploying source b: The dynamics of x_0, x_{a_1}, x_{a_2}, x_{b_1} and x_{b_2} over time where $\gamma_a = 1$. (Initial fraction of nodes in different states: $x_0(0) = 0.98$, $x_{a_2}(0) = 0.02$ and $x_{b_2}(0) = 0.01$.) (a) $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 16$, $\lambda = 2$, $\gamma_b = 1$. (b) $\alpha_1 = \alpha_2 = 16$, $\beta_1 = \beta_2 = 21$, $\lambda = 2$, $\gamma_b = 1$. (c) $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 16$, $\lambda = 2$, $\gamma_b = 0.5$. (d) $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 16$, $\lambda = 31$, $\gamma_b = 1$.

by dominant source b). For the ease of presentation, we only show the fraction of nodes in state 0, state $a_{\mathcal{M}}$ and state $b_{\mathcal{N}}$.

We carried out a set of experiments where $\mathcal{M} = 3$ and $\mathcal{N} = 2$. Figure 16 shows the numerical results. Figure 16(b) shows that even if $\alpha_{1,2,3}$ is larger than $\beta_{1,2}$ and $x_{a_3}(0) = 2x_{b_2}(0)$, source a may still die out. For an application of this model, assume source a and b are two products. If product b enters the market later than product a, or customers who brought product b cannot contact as many potential customers as those who brought product a, decreasing the phase for product b is *crucial*. Figure 16(c) and (d) show that if source a takes more phases to activate nodes, it needs to have a larger transmission rate so as to counter the dominant effect of b.

6. Related work

Contagion modelling on large-scale networks has gained a lot of attention lately. It is important for researchers not only to gain the fundamental insights on how the disease, idea or behaviour spreads, but also how products get promoted in social networks. One popular epidemic model is the SIS model [14–16]. A series of works focus on the analysis of the SIS model in different networks [2,7,17–20]. For epidemic models, the study of epidemic thresholds has received significant attention. Ganesh *et al.* [2] and Chakrabarti *et al.* [7] analysed the epidemic thresholds for the SIS model on general undirected
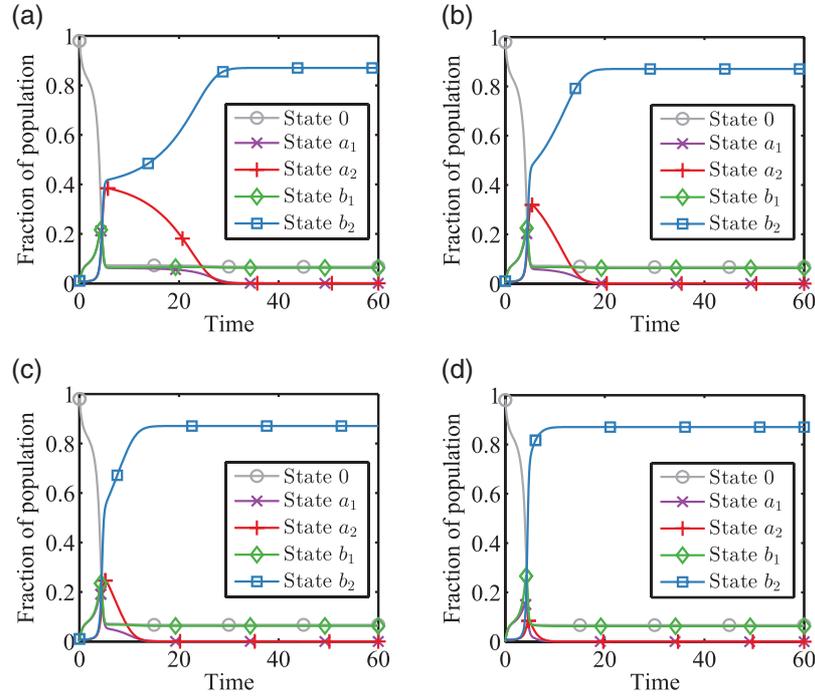
FIG. 15. Numerical results over a complete graph showing the impact of λ: The dynamics of x_0, x_{a_1}, x_{a_2}, x_{b_1} and x_{b_2} over time where $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 16$ and $\gamma_a = \gamma_b = 1$. (Initial fraction of nodes in different states: $x_0(0) = 0.97$, $x_{a_2}(0) = 0.01$ and $x_{b_2}(0) = 0.01$.) (a) $\lambda = 0.1$. (b) $\lambda = 0.5$. (c) $\lambda = 1$. (d) $\lambda = 3$.

networks. Recently, Prakash *et al.* [3] gave the threshold for a series of virus propagation models, which include the SIS model, on arbitrary graphs. Other well-known models, such as SEIR and SEIV model, that introduce the "exposed" state have also been proposed.

In practice, however, a more general SIS model is needed. In particular, one 'exposed' state between healthy and infected states may not be enough for modelling disease spreading. For viral marketing, the consumer purchase decision process theory [6] suggests that there are five stages (or states) until a consumer buys a product and influences others. This motivates us to study and analyse a generalized SIS model that allows multi-susceptible states before getting infected. To the best of our knowledge, previous work cannot be easily extended on our generalized multi-state SIS model. We use the mean-field analysis technique to analyse our generalized model, and show that our methodology predicts the diffusion accurately for many different types of graphs.

Recently, there has been a research trend on modelling and analysing competing processes. Melnik *et al.* [21] proposed a model of a multi-stage complex contagion, in which agents at different stages exert different amounts of influence on their neighbours. Our work focused on the generalized SIS model, in which the phase-transition process is different from the cascade model used in [21]. Newman [22], Aspnes *et al.* [23], Beutel *et al.* [24] and Prakash *et al.* [8] studied the scenario where two sources are competing, but their models are defined differently from ours since they did not consider intermediate stages between susceptible and infected states.
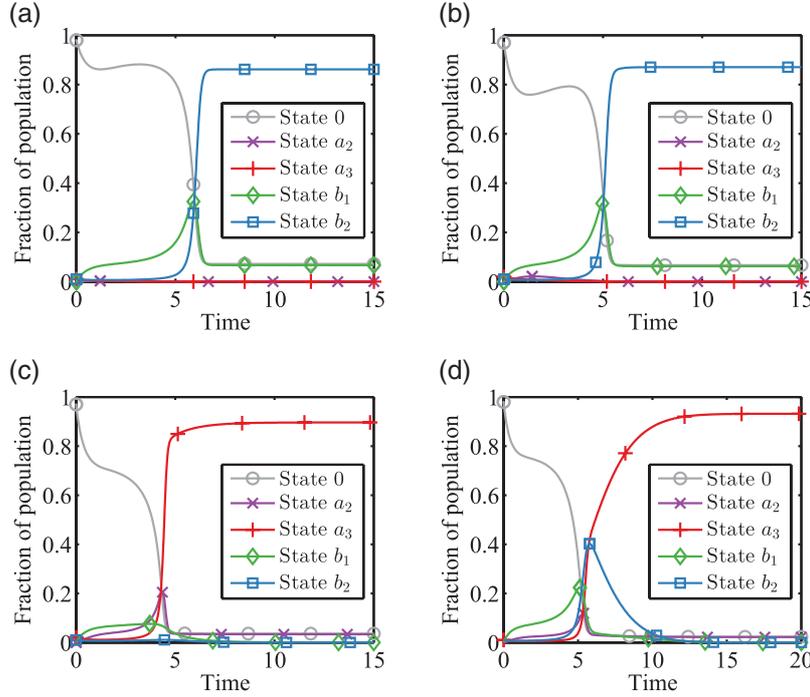
FIG. 16. Numerical results over a complete graph showing the impact of $\mathcal{M}$ and $\mathcal{N}$: The dynamics of x_0, $x_{a_{\mathcal{M}-1}}$, $x_{a_{\mathcal{M}}}$, $x_{b_{\mathcal{N}-1}}$ and $x_{b_{\mathcal{N}}}$ over time where $\mathcal{M} = 3$, $\mathcal{N} = 2$, $\gamma_a = \gamma_b = 1$ and $\lambda_i = 2$ ($1 \leqslant i \leqslant \mathcal{M} - 1$). (Initial fraction of nodes in different states: $x_{b_2}(0) = 0.01$.) (a) $\alpha_i = 25, \beta_1 = \beta_2 = 16, x_{a_3}(0) = 0.02$. (b) $\alpha_i = 30, \beta_1 = \beta_2 = 16, x_{a_3}(0) = 0.02$. (c) $\alpha_i = 20, \beta_1 = \beta_2 = 15, x_{a_3}(0) = 0.01$. (d) $\alpha_i = 45, \beta_1 = \beta_2 = 16, x_{a_3}(0) = 0.01$.

In our early work [25], we presented the idea of the generalized SIS model and proposed a generalized SIS model with two competing sources. For a generalized SIS model, our analysis focused on complete graphs in [25] and we extend our analysis to general graphs in this paper.

7. Conclusion

In this paper, we propose a generalized SIS model by allowing the existence of intermediate states between susceptible and infected states. We use the *mean-field analytical technique* to analyse the influence spreading dynamics in both complete graphs and general graphs. Specifically, we study the problem of determining the sufficient condition that prevents the information or virus spreading, or equivalently, the necessary condition that may lead to the information or virus spreading. We show that the condition that prevents the spread of contagions depends on two de-coupled effects: the network topology and the parametric values of the generalized SIS model. For undirected graphs, the condition depends on the largest eigenvalue of the adjacency matrix, the infection rates, the recovery rate and the number of states k in our generalized SIS model. For directed graphs, besides the above factors, the condition also depends on the maximum incident degree among all nodes. We carry out extensive experiments on both synthetic and large real datasets to show the dynamics on fractions of states in the general SIS model with different parametric values operating on different networks, and illustrate our conditions for the tipping point.

We also analyse two competing sources, one dominant and one regressive, under the generalized SIS model which allows multi-intermediate states. In particular, we allow nodes being exposed to the regressive source change to being influenced by the dominant source. We formulate the dynamic process on both complete and general graphs and show via experiment how different parametric values, such as different transmission rates or initial fraction of nodes being infected, may affect the phase transition results and final equilibrium.

We believe our work is a step towards building more realistic models to describe the spread of information or virus. We like to note that our models in this paper, especially the generalized SIS model with two competing sources, leave important open questions for future research. For example, is it possible to predict the final fraction of nodes being infected quickly given any network topology and parametric values of the model? Moreover, for the general SIS model with two or more competing sources, it is interesting to find the condition that leads to the spread of the dominant source or prevents the spreading of any sources. We will leave these questions as future work.

Funding

References

1. NEWMAN, M. E. (2002) Spread of epidemic disease on networks. *Phys. Rev. E*, **66**, 016128.
2. GANESH, A., MASSOULIÉ, L. & TOWSLEY, D. (2005) The Effect of Network Topology on the Spread of Epidemics. *Proceedings IEEE INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2. Piscataway, USA: IEEE, pp. 1455–1466.
3. PRAKASH, B. A., CHAKRABARTI, D., VALLER, N. C., FALOUTSOS, M. & FALOUTSOS, C. (2012) Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowl. Inf. Syst.*, **33**, 549–575.
4. LI, Y., ZHAO, B. Q. & LUI, J. (2012) On modeling product advertisement in large-scale online social networks. *IEEE/ACM Trans. Netw.*, **20**, 1412–1425.
5. BHAGAT, S., GOYAL, A. & LAKSHMANAN, L. V. (2012) Maximizing product adoption in social networks. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*. New York, USA: ACM, pp. 603–612.
6. BLACKWELL, R., MINIARD, P. & ENGEL, J. (2006) *Consumer Behavior*. Mason, USA: Thomson/South-Western.
7. CHAKRABARTI, D., WANG, Y., WANG, C., LESKOVEC, J. & FALOUTSOS, C. (2008) Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, **10**, 1.
8. PRAKASH, B. A., BEUTEL, A., ROSENFELD, R. & FALOUTSOS, C. (2012a) Winner takes all: competing viruses or ideas on fair-play networks. *Proceedings of the 21st International Conference on World Wide Web*. New York, USA: ACM, pp. 1037–1046.
9. CHUNG, F. & LU, L. (2002) Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, **6**, 125–145.
10. OPSAHL, T. & PANZARASA, P. (2009) Clustering in weighted networks. *Soc. Netw.*, **31**, 155–163.
11. KLIMT, B. & YANG, Y. (2004) The enron corpus: A new dataset for email classification research. *Machine Learning: ECML 2004*. Berlin: Springer, pp. 217–226.
12. LESKOVEC, J., LANG, K. J., DASGUPTA, A. & MAHONEY, M. W. (2009) Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math.*, **6**, 29–123.

13. RICHARDSON, M., AGRAWAL, R. & DOMINGOS, P. (2003) Trust management for the semantic web. *The Semantic Web-ISWC 2003*. Berlin: Springer, pp. 351–368.

14. ANDERSON, R. M. & MAY, R. M. (1991) *Infectious Diseases of Humans*, vol. 1. Oxford: Oxford University Press.

15. BAILEY, N. T. (1975) *The Mathematical Theory of Infectious Diseases and its Applications*. 5a Crendon Street, High Wycombe, Bucks HP13 6LE: Charles Griffin & Company Ltd.

16. DALEY, D. J., GANI, J. & GANI, J. J. M. (2001) *Epidemic Modelling: An Introduction*, vol. 15. Cambridge: Cambridge University Press.

17. EGUILUZ, V. M. & KLEMM, K. (2002) Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.*, **89**, 108701.

18. GROSS, T., D'LIMA, C. J. D. & BLASIUS, B. (2006) Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.*, **96**(20), 208701.

19. LÓPEZ-PINTADO, D. (2008) Diffusion in complex social networks. *Games Econom. Behav.*, **62**, 573–590.

20. PARSHANI, R., CARMI, S. & HAVLIN, S. (2010) Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Phys. Rev. Lett.*, **104**, 258701.

21. MELNIK, S., WARD, J. A., GLEESON, J. P. & PORTER, M. A. (2013) Multi-stage complex contagions. *Chaos*, **23**, 013124.

22. NEWMAN, M. E. (2005) Threshold effects for two pathogens spreading on a network. *Phys. Rev. Lett.*, **95**, 108701.

23. ASPNES, J., RUSTAGI, N. & SAIA, J. (2007) Worm versus alert: Who wins in a battle for control of a large-scale network? *Principles of Distributed Systems*. Berlin: Springer, pp. 443–456.

24. BEUTEL, A., PRAKASH, B. A., ROSENFELD, R. & FALOUTSOS, C. (2012) Interacting viruses in networks: can both survive? *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: ACM, pp. 426–434.

25. LIN, Y., LUI, J., JUNG, K. & LIM, S. (2013) Modeling multi-state diffusion process in complex networks: theory and applications. *2013 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. Piscataway, USA: IEEE, pp. 506–513.