Physica A 471 (2017) 325–336

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Evaluation of an analytic, approximate formula for the time-varying SIS prevalence in different networks

Qiang Liu *, Piet Van Mieghem

Delft University of Technology, Faculty of EECS, P.O. Box 5031, 2600 GA Delft, The Netherlands

HIGHLIGHTS

- An approximate formula for the time-dependent SIS prevalence is evaluated.
- The formula is based on spectral decomposition of the Laplacian matrix of networks.
- The formula is comparable with mean-field approximations.
- The Riccati equation approximates the Markovian SIS prevalence well.

ARTICLE INFO

Article history:
Received 6 October 2016
Received in revised form 15 November 2016
Available online 24 December 2016

Keywords:
SIS epidemic process

ABSTRACT

One of the most important quantities of the exact Markovian SIS epidemic process is the time-dependent prevalence, which is the average fraction of infected nodes. Unfortunately, the Markovian SIS epidemic model features an exponentially increasing computational complexity with growing network size N. In this paper, we evaluate a recently proposed analytic approximate prevalence function introduced in Van Mieghem (2016). We compare the approximate function with the N-Intertwined Mean-Field Approximation (NIMFA) and with simulation of the Markovian SIS epidemic process. The results show that the new analytic prevalence function is comparable with other approximate methods.

1. Introduction

Many real-world spreading phenomena can be described by an epidemic process on a network, such as the spreading of diseases, of information and rumors, failures of computers in a network. The prevalence $y(t)$, which is the average fraction of infected nodes in a network at time t, contains a wealth of information about the spreading process on a graph. Given the initial condition and the network topology, the prevalence can be determined at each time. However, the prevalence is not easy to calculate, because the topology is complex for large scale networks. Mean-field approximations are usually introduced to allow a feasible computation.

The Susceptible–Infected–Susceptible (SIS) epidemic model is a basic model of epidemiology [1,2]. The network topology is represented by a graph constructed by nodes and links. For each node, there are two states: a susceptible (or healthy) and an infected state. Each infected node can infect its healthy neighbors, with an infection rate β. Each infected node can also be cured at a constant rate δ. There exists an epidemic threshold τ_c, such that, when the effective infection rate $\tau = \beta/\delta \geq \tau_c$, the virus spreads, and the prevalence stays on a steady level for a long time with a dynamic balance between infection and curing. If $\tau < \tau_c$, the virus will extinct from the network exponentially fast for sufficiently long time. However, the exact

* Corresponding author.
E-mail addresses: Q.L.Liu@tudelft.nl (Q. Liu), P.F.A.VanMieghem@tudelft.nl (P. Van Mieghem).

http://dx.doi.org/10.1016/j.physa.2016.12.031
0378-4371/

SIS epidemic threshold of general graphs is unfortunately unknown. Some approximate thresholds have been proposed [3–5], such as the lower bound $\tau_c^{(1)} = 1/\lambda_1$, where λ_1 is the largest eigenvalue (spectral radius) of the adjacency matrix A of the graph, and the heterogeneous mean-field threshold $\tau_c^{\mathrm{HMF}} = E[D]/E[D^2]$, where D is the degree of a randomly chosen node. However, the accuracy of these approximate thresholds and also the accuracy of the mean-field prevalence varies over networks. In this paper, we compare the exact prevalence, the N-Intertwined Mean-Field Approximation (NIMFA) and a new analytical prevalence function introduced in [6].

The paper is organized as follows. In Section 2, we briefly introduce the new analytical approximate function for the prevalence proposed in [6]. In Section 3, we compare the prevalence function with simulation results and NIMFA on the Erdős–Rényi graph, the cycle graph, the star graph, and the Barabási–Albert power-law graph. In Section 4, we further investigate the approximation of the remainder Ψ by higher degree polynomials rather than constants. Finally, we conclude the paper. Background on SIS, NIMFA and the simulation are placed in the Appendix.

2. Tanh-formula: an analytical approximate function for time-dependent analysis of prevalence

A Bernoulli random variable $X_j(t) \in \{0, 1\}$ indicates the infection state of node j at time t: $X_j(t) = 1$ means that the node j is infected at time t, while $X_j(t) = 0$ refers to a healthy or cured node j at time t. The Laplacian matrix of a graph are denoted by Q.

The prevalence of the exact Markovian SIS epidemic process obeys the equation [7],

$$\frac{dy(t)}{dt} = -y(t) + \frac{\tau}{N} E\left[w(t)^T Q\, w(t)\right] \tag{1}$$

where $w(t) = [X_1(t), \ldots, X_N(t)]$ is the network state vector. From [8, Art. 77], the quadratic Laplacian $w(t)^T Q\, w(t)$ in (1) equals the number of the links which have only one end node infected at time t. The Eq. (1) has a clear physical meaning that the changing rate of prevalence y is equal to the difference between the expectation of the number of links with one end node infected $E[w^T Q w]$ and the prevalence y. More generally, the exact governing Eq. (1) also applies in temporal networks [9], where the connections between nodes change with time and the Laplacian $Q(t)$ is time-varying. After spectral decomposition of $E[w^T Q w]$ in [6], the governing equation is,

$$\frac{dy(t)}{dt} = (\tau\mu_{N-1} - 1)y(t) - \tau\mu_{N-1}y^2(t) - \Psi \tag{2}$$

where μ_{N-1} is the algebraic connectivity, the second smallest eigenvalue of the Laplacian matrix Q and the parameter Ψ is a time-dependent remainder derived in [6]. In [6], the remainder Ψ is bounded by two constants and (2) reduces to a Riccati differential equation that can be solved explicitly. By assuming that the remainder Ψ is constant and equal to c, the solution of (2) is

$$\tilde{y}(t) = \frac{1}{2}\left(1 - \frac{1}{\tau\mu_{N-1}}\right) + \frac{\Xi}{2}\tanh\left(\frac{\tau\mu_{N-1}\Xi}{2}t + \operatorname{arctanh}\left(\frac{2y_0 - \left(1 - \frac{1}{\tau\mu_{N-1}}\right)}{\Xi}\right)\right) \tag{3}$$

where $\Xi = \sqrt{(1 - \frac{1}{\tau\mu_{N-1}})^2 - \frac{4c}{\tau\mu_{N-1}}}$ and $y_0 = y(0)$. Eq. (3) is called the tanh-formula, which describes the time-dependent prevalence of epidemic process on networks. Moreover, the tanh-formula (3) only depends on the algebraic connectivity μ_{N-1} of the network apart from the constant c and the initial condition y_0. For a same initial condition [6], the prevalence is bounded by the explicit analytic function of time $\tilde{y}(t)|_{c=c_U} \le y(t) \le \tilde{y}(t)|_{c=c_L}$, when the remainder Ψ is bounded by $c_L \le \Psi \le c_U$. The tanh-formula (3) is the result of an approximation of the spectral decomposition of the Laplacian Q and properties of the Bernoulli vector w, which is entirely different from a mean field concept.

The remainder Ψ is time-dependent, which makes it hard to find a suitable constant c for Eq. (3). Due to the absorbing state in a SIS Markovian process, the steady-state of the finite-state SIS model is the trivial overall-healthy state. A possible way to describe the SIS metastability is by adding a self-infection process, where every healthy node can be self-infected with rate ϵ, which is the so-called ϵ-SIS model [10]. Without an absorbing state, the ϵ-SIS Markov chain has a unique steady state, which can be used to approximate the metastable state of the classic SIS model ($\epsilon = 0$). Another way is to remove the all-healthy state preventing the virus from dying out [11]. However, the drawback of the above-mentioned approximations for the metastable state is that the steady state in the modified models is independent of the initial conditions, but the metastable state depends on the initial conditions [12]. If we assume that the prevalence is approximately equal to y_m in the metastable state, where $dy/dt \approx 0$, the remainder Ψ in the metastable state is

$$\Psi_m \approx (\tau\mu_{N-1} - 1)y_m - \tau\mu_{N-1}(y_m)^2. \tag{4}$$

A suitable $c = \Psi_m$ for (3) can be selected since the metastable state lasts long, and the remainder Ψ in the metastable state is almost constant. Although formula (3) assumes that parameter c is independent of the prevalence y, we will present simulation results, which show that when $c = \Psi_m$, the tanh-formula (3) performs exactly in the metastable state and also well in the initial transition regime.

For a fixed network with constant algebraic connectivity μ_{N-1} and effective infection rate τ, the time-dependent prevalence $y(t)$ of the exact Markovian SIS epidemic model should only be determined by the initial condition. If we choose $c = \Psi_m$ in the tanh-formula (3), then the prevalence $\tilde{y}(t)$ is determined both by the initial condition y_0 and the prevalence in the metastable state y_m. In (3), only the algebraic connectivity μ_{N-1} appears, and other information of the network is condensed, via Ψ, into the constant c.

3. Comparison of tanh-formula, NIMFA, and simulation

We present simulation results on four basic graph models: the Erdős–Rényi (ER) random graph, the star graph, the cycle graph, and the Barabási–Albert power-law graph [13]. The Erdős–Rényi random graph $G_p(N)$ has N nodes and is constructed by independently connecting two nodes with probability p. The star graph with N nodes has only one center node connected to $N-1$ leaf nodes, but leaf nodes do not connect to each other. The cycle graph is a graph where nodes are placed in a cycle and only two nearest neighbor nodes on that cycle are connected. In the comparison, the network size of the three above-mentioned kinds of graphs is $N = 50$, because the prevalence is more sensitive to the change of the number of infected nodes, and is better observed on small graphs. In the power-law graph, the network size is relatively larger with $N = 1000$ in order for the degree distribution to be clearly power law [14]. The simulation method is described in Appendix C.

The tanh-formula approximation $\tilde{y}(t)$ in (3) is compared with the NIMFA prevalence $y^{(1)}(t)$ in (B.5) and the simulation results. The constant c in the tanh-formula (3) is selected as the remainder Ψ_m of the metastable state from the simulation. In each realization of the simulation, the same number of infected nodes are picked up randomly at $t = 0$ and the fraction of infected nodes in the epidemic processes is sampled at the same time points (the sampling rate is 100 points per normalized time unit[1]). Then, the simulated prevalence is obtained by averaging the sampled values over all the 10^6 realizations at each time point.

In the following, we compare the tanh-formula with NIMFA and simulation results under three different regimes: $\tau > \tau_c$, $\tau \approx \tau_c$ and $\tau < \tau_c$.

3.1. ER random graph

In Fig. 1(a) where the normalized effective infection rate is $x = \tau/\tau_c^{(1)} = \lambda_1\tau = 3.12$ for the ER random graph, the effective infection rate τ is above the threshold τ_c, which means that the virus can spread over the entire network. If the virus is able to spread on the network, the epidemic will stay in a metastable state for a very long time [1,15]. The prevalence in the metastable state is determined both by the effective infection rate τ and the initial fraction of infected nodes. Moreover, the time for an epidemic process to reach the metastable state also varies with different initial conditions [16]. Fig. 1(a) shows that for a same effective infection rate τ, even only one additional initially infected node can make a significant increase in the prevalence in the metastable state. The main reason for this difference is that, there is a non-negligible die-out probability when the number of initially infected nodes is small [12]. When all nodes are infected initially, the prevalence in the metastable state reaches to a maximum and is very well estimated by the upper bound NIMFA.

Fig. 1(b) shows similar results with a higher normalized effective infection rate $x = \tau/\tau_c^{(1)} = 5.21$. With a higher effective infection rate τ, NIMFA performs better compared to Fig. 1(a). When all nodes are infected initially, the NIMFA prevalence is almost the same as the simulated prevalence. The tanh-formula approximation (3) performs well, when the effective infection rate τ is high.

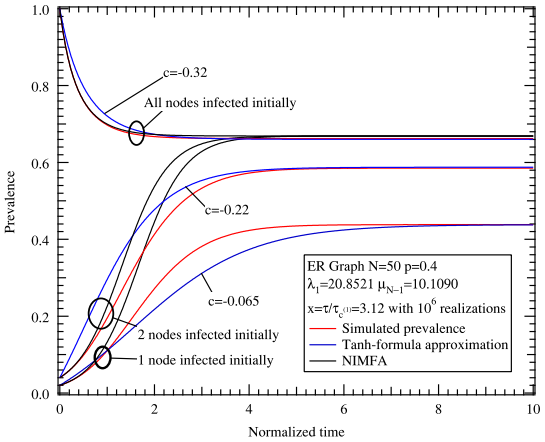
Fig. 2 illustrates the prevalence with the normalized effective infection rate $x = 0.83$. The prevalence decreases exponentially fast for sufficiently large time [17,6]. On a different scale, the curves are similar and NIMFA upper bounds the prevalence. When the effective infection rates $\tau < \tau_c$, the constant c in tanh-formula (3) always equals 0 as deduced from (4). Both NIMFA and the tanh-formula perform well.

For the condition that the effective infection rate τ is around $\tau_c^{(1)}$, the change of prevalence with time t is more complex. Since the NIMFA threshold $\tau_c^{(1)} = 1/\lambda_1$ is a lower bound of the actual threshold τ_c, when $\tau_c^{(1)} \le \tau \le \tau_c$, NIMFA fails to estimate the exponentially decreasing prevalence. Indeed, Fig. 3(a) shows that the NIMFA prevalence (B.5) converges to a non-zero value, but the exact prevalence decreases exponentially fast to zero. In Fig. 3(b), the effective infection rate τ is around the actual threshold and the prevalence experiences an increasing period and then decreases to 0 slowly. Though there does not exist an obvious metastable state, the virus can survive for a relatively long time period. The tanh-formula (3) also fails to estimate the exact prevalence when τ is around τ_c, since the analytical approximation $\tilde{y}(t)$ is always monotonically increasing or decreasing.

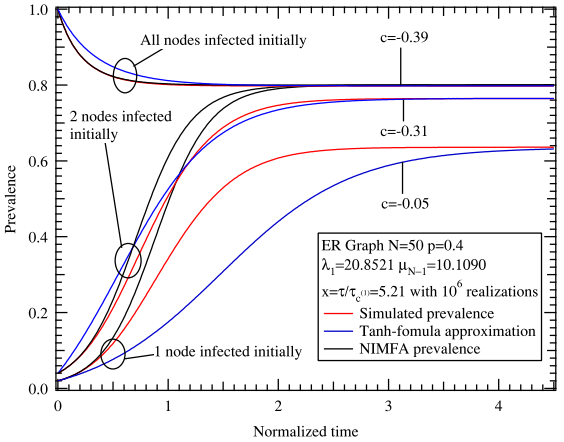
3.2. Cycle graph

On cycle graphs, the diameter, which is the longest shortest path, is $N/2 - 1$ for N even or $(N-1)/2 - 1$ for N odd. The diameter of the cycle graph is much larger than that of a general ER random graph. Generally, the virus spreads slower in a

[1] A normalized time unit is defined when the curing rate $\delta = 1$. Physically, one node is cured per time unit on average.



(a) The normalized effective infection rate $x = \tau/\tau_c^{(1)} = 3.12$.



(b) The normalized effective infection rate $x = 5.21$.

Fig. 1. Results on the ER graph $G_{0.4}(50)$ when the effective infection rate τ is above the epidemic threshold τ_c. The largest eigenvalue of adjacency matrix A is $\lambda_1 = 20.8581$ and the algebraic connectivity is $\mu_{N-1} = 10.1090$. The prevalence is the average fraction of infected nodes of 10^6 realizations.

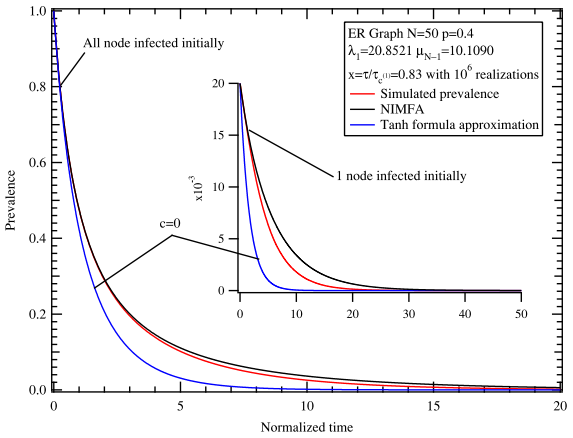
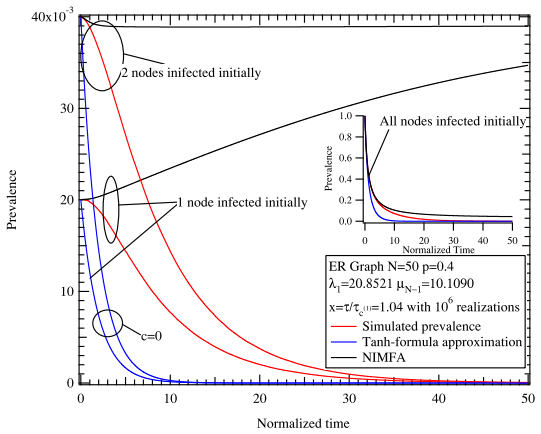
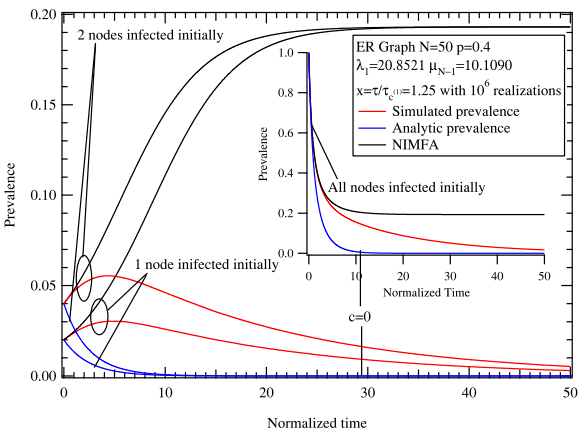


Fig. 2. Results on the same ER graph $G_{0.4}(50)$. The normalized effective infection rate is $x = 0.83$ that the infection rate τ is below the threshold τ_c.



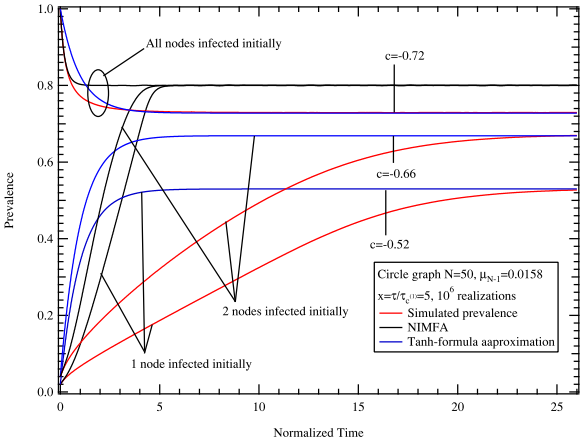
(a) The normalized infection rate $x = 1.04$.



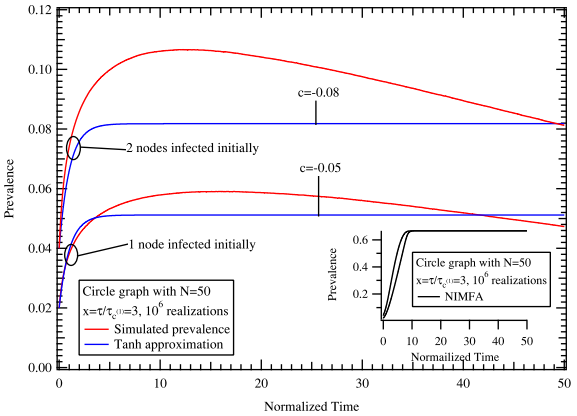
(b) The normalized effective infection rate is $x = 1.25$.

Fig. 3. Results on the graph $G_{0.4}(50)$ when the effective infection rate τ is around the epidemic threshold.

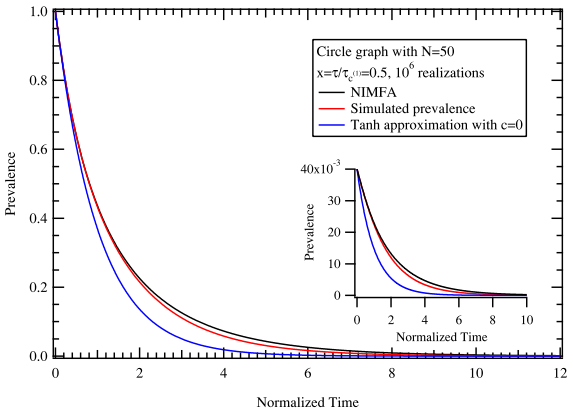
graph with longer diameter. In Fig. 4(a), where the normalized effective infection rate $x = 5$ is very large, the time for the network to reach the metastable state is still long with only a few nodes infected initially. The approximate prevalence of NIMFA and the tanh-formula increase faster to metastable state prevalence than the simulated prevalence. Similar to the ER



(a) The normalized effective infection rate $x = 5$, above the epidemic threshold.



(b) The normalized effective infection rate $x = 3$, around the threshold.



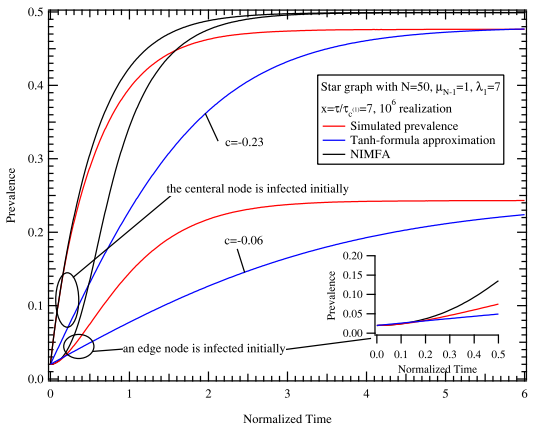
(c) The normalized effective infection rate $x = 0.5$, below the threshold.

Fig. 4. Simulation and analytic results of the SIS epidemic process on the cycle graph with $N = 50$ and $\delta = 1$. The largest eigenvalue of adjacency matrix A is $\lambda_1 = 2$ and the algebraic connectivity is $\mu_{N-1} = 0.0158$.

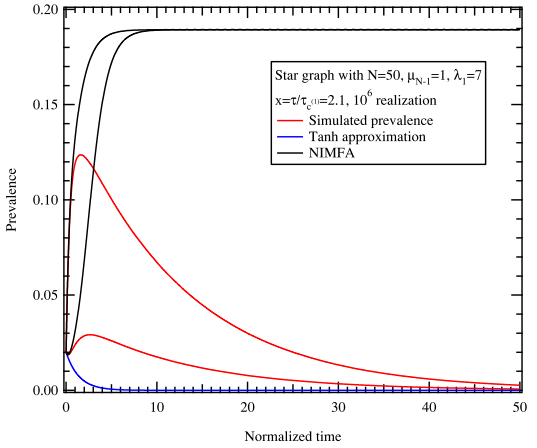
random graph, the addition of only one initially infected node increases the metastable prevalence substantially. However, adding another infected node does not obviously shorten the time, which is needed by the epidemic process to reach the metastable state. Furthermore, different from the ER random graph, there is still a gap between the NIMFA prevalence and the simulated prevalence when all nodes infected initially. The epidemic threshold of cycle graphs [18] is about $\tau_c = 1.65$ ($x = 3.3$) and the effective infection rate τ in Fig. 4(b) is around τ_c. NIMFA still fails to estimate the exact prevalence for $\tau \approx \tau_c$. Additionally, similar to the same condition that $\tau \approx \tau_c$ in ER random graphs, the tanh-formula approximation fails to describe the non-monotonical prevalence. For the cycle graph, both the tanh-formula and NIMFA are not good enough to estimate the time-dependent prevalence even with a high effective infection rate τ. Only when $\tau < \tau_c$, the analytic tanh-formula and NIMFA are both working well as shown in Fig. 4(c).

3.3. Star graph

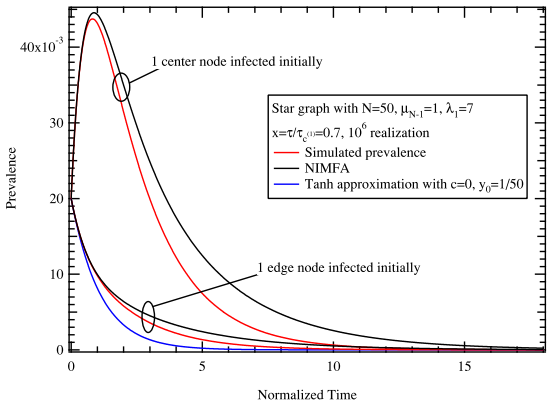
On star graphs, the location of the initially infected node is critical. The transition rate [11] from the state, where the center node is infected, to the states that the center node and an arbitrary leaf node are infected, is $(N-1)\beta$. The curing rate is δ, which means that the transition rate from the state that the center node is infected to the state that all nodes are cured (virus dies out) is δ. Given the two transition rates, the die-out probability at the first state transition is $\delta/((N-1)\beta+\delta) = 1/(\tau(N-1)+1)$, which decreases linearly with network size N. If a leaf node is infected initially, the die-out probability at the first state transition is a relatively larger constant $\delta/(\delta+\beta) = 1/(\tau+1)$, where β is the rate of infecting the center node. The virus needs to survive the dangerous period at the initial stage and after the central node is infected, the virus begins to spread to the whole network. Fig. 5(a) shows the spreading delay at the initial stage when only a leaf node is infected initially. The relative bigger die-out probability at the initial stage causes a gap in the prevalence in the metastable state between the two different initial conditions that the center node or one leaf node is infected. Contrary to the cycle graphs, star graphs have a very small diameter (only 2). The network reaches the metastable state very quickly and earlier



(a) The normalized effective infection rate $x = 7$, above the threshold.



(b) The normalized effective infection rate is $x = 2.1$, around the threshold.



(c) The normalized effective infection rate $x = 0.7$, below the threshold.

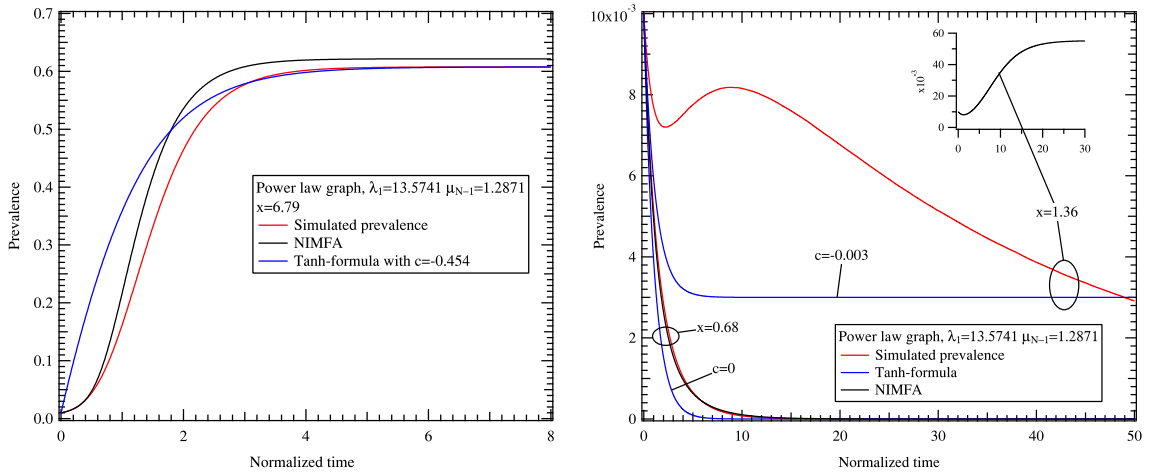
Fig. 5. Simulation and analytic results of the SIS epidemic process on the star graph with $N = 50$ and $\delta = 1$. The largest eigenvalue of adjacency matrix A is $\lambda_1 = 7$ and the algebraic connectivity is $\mu_{N-1} = 1$.

than the tanh-formula approximation. For an epidemic process with a lower effective infection rate τ around the epidemic threshold τ_c, as shown in Fig. 5(b), both NIMFA and the tanh-formula approximation fail to estimate the prevalence. Fig. 5(c) shows the results of $\tau < \tau_c^{(1)}$ where the effective infection rate τ is below the threshold, which verifies the conclusion in [6] that the prevalence can still increase at the initial stage, even if the effective infection rate τ is below the threshold. The NIMFA prevalence $y^{(1)}(t)$ shows the same trend as the prevalence $y(t)$. Fig. 5(a), (b) and (c) illustrate that, if the virus spreads from the center node, NIMFA gives a very good prediction for the initial increase (but not good in the long term when the effective infection rate $\tau \approx \tau_c$), irrespective of the effective infection rate τ. The tanh-formula (3) only works well when τ is higher or lower than the threshold with a leaf node infected initially.

3.4. Barabási–Albert power-law graph

The power-law network is generated start from 3 initial nodes, and each time a new node with 3 edges is added to the graph. The newly added node is connected to an existing node j with probability $d_j/2L$, where d_j is the degree of node j and L is the current number of the edges. Fig. 6(a) and (b) show the comparison of the three methods on a power-law graph with network size $N = 1000$. Also, the prevalence is obtained by averaging 10^6 realizations of the simulation. When the infection rate is above the threshold with $x = 6.79$, both the tanh-formula (3) and NIMFA show an accurate approximation to the prevalence. With the infection rate $x = 0.68$ below the threshold, the prevalence decreases exponentially fast to 0, and both the tanh-formula (3) and NIMFA are also accurate as shown in Fig. 6(b). However, when the infection rate $x = 1.36$ is around the threshold as shown in Fig. 6(b), both the tanh-formula and NIMFA fail to estimate the prevalence.

In the power-law graph, we find that NIMFA is accurate at the very short initial stage, which is similar to the results in the star graph when the center node is infected initially. Nevertheless, the accuracy of tanh-formula (3) is comparable with NIMFA in the long run.



(a) The normalized effective infection rate $x = 6.79$, above the threshold.

(b) The infection rates are around and below the threshold.

Fig. 6. Results on the Barabási–Albert power-law graph $N = 1000$.

4. Approximating the remainder Ψ with polynomials of higher degree

In this section, the remainder Ψ in the Eq. (2) and the constant c in (3) will be further discussed. As mentioned above, the remainder Ψ is considered as a constant to derive the tanh-formula (3), and we have chosen (4) in this paper. Thus, the difference $\Psi - c$ determines the accuracy of the tanh-formula (3), especially in the transition regime of the prevalence. Fig. 7 presents the time-dependent remainder Ψ with one node infected initially, for effective infection rates above the thresholds in an ER random graph $G_{0.4}(50)$ (the same graph as in Section 3.1). Fig. 7 indicates that the time-dependent Ψ has an undershoot in the transition regime of prevalence. Comparing Fig. 7 with Fig. 1(a) and (b), a larger undershoot of the remainder Ψ in the transition regime leads to a less accurate tanh-formula (3). Furthermore, the length of the time interval of the undershoot, which is also the length of the transition regime of the prevalence, decreases with the increase of the normalized effective infection rate x, while the amplitude of the undershoot increases with x. From [6], the quadratic Laplacian is derived as $w^T Q w = \mu_{N-1} N(S - S^2) + R$, where the parameter R is a correction and also explicitly expressed in [6]. Only in complete graphs where $\mu_{N-1} = N$ and $E[R] = 0$, the remainder Ψ has a physical meaning $\Psi(t) = \tau N \mathrm{Var}[S(t)]$, where the variance of the fraction S of infected nodes is prominent. The remainder Ψ is hard to handle because Ψ contains both the variance of the number of infected nodes $\mathrm{Var}[S]$ and a correction R, where R is related to the projections [6] of the network state w on the eigenvectors of the Laplacian Q. An equation of $\mathrm{Var}[S]$ is possible to obtain, but higher order moments will be introduced [6]. Apart from taking Ψ as constant in (2), another possible way is to approximately express Ψ as a function of the prevalence y: $\Psi(t) \approx f(y(t))$. We apply the minimum Chi-square method to fit the curves in this paper [19]. In the ER random graph with a high effective infection rate $\tau > \tau_c$, we find that the remainder Ψ can be fitted very well by quartic polynomials $\Psi(t) \approx a y^4(t) + b y^3(t) + c y^2(t) + d y(t) + e$, where a, b, c, d, e are constant coefficients. Thus, a closed non-linear governing equation can be obtained as

$$\frac{dy}{dt} = (\tau \mu_{N-1} - 1) y - \tau \mu_{N-1} y^2 - (a y^4 + b y^3 + c y^2 + d y + e). \tag{5}$$

Moreover, a quadratic fitting polynomial of the remainder $\Psi \approx f y^2 + g y + h$ leads to an analytically solvable Riccati differential equation

$$\frac{dy}{dt} = (\tau \mu_{N-1} - 1 - g) y - (\tau \mu_{N-1} + f) y^2 - h. \tag{6}$$

Here, we present an example of expressing the remainder Ψ as a function of the prevalence y. Fig. 8(a), (b) and (c), which are correspond to Fig. 1(a), illustrate the fitting results of the remainder $\Psi(y)$ with quartic, cubic and quadratic polynomials, respectively. When all nodes are infected initially, the remainder $\Psi(y)$ is fitted well with a quadratic polynomial; when only a few nodes infected initially, higher order polynomials are better to fit the remainder $\Psi(y)$. Fig. 8(d) shows the comparison of the simulated prevalence and the solutions of the approximated governing equations in form (5), where the remainder Ψ is fitted with polynomials. All the approximate governing equations describe the prevalence well. Thus, a Riccati governing equation as in (6) is simpler to describe an epidemic process if we know the coefficients f, g, h of the quadratic polynomial $f y^2 + g y + h$.

However, substituting the remainder Ψ with polynomials is not always possible. As illustrated in Fig. 9, the effective infection rate τ is around the threshold τ_c. There are multiple values of remainder Ψ mapping to a same certain value of the

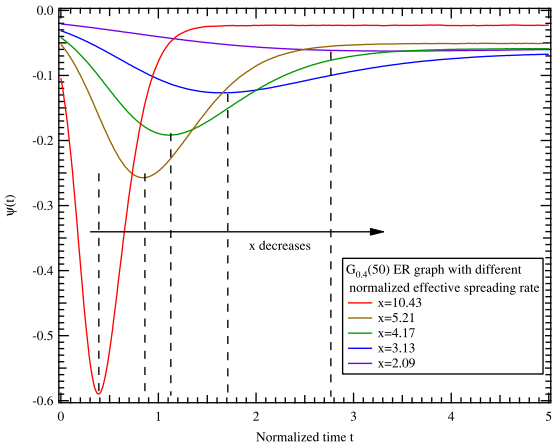
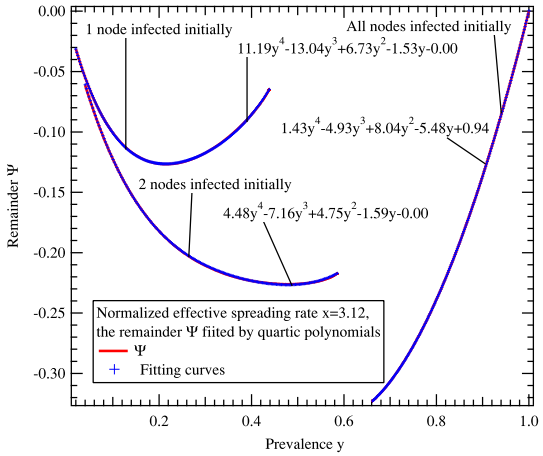
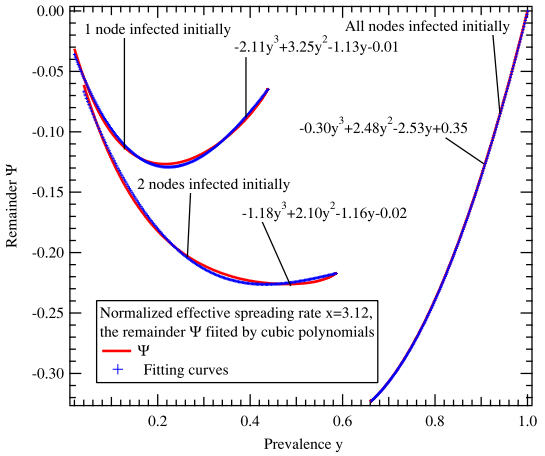


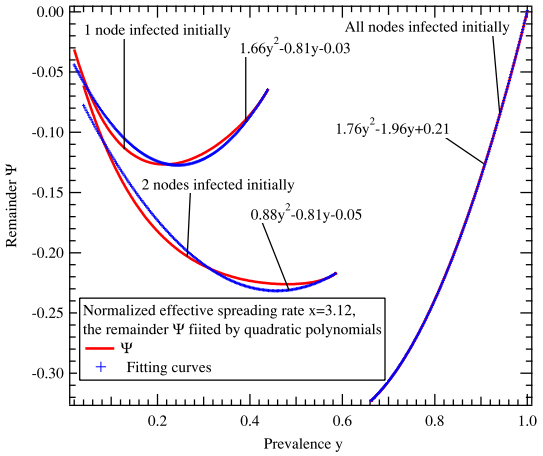
Fig. 7. Time-dependent Ψ: the transition regime and the undershoots.



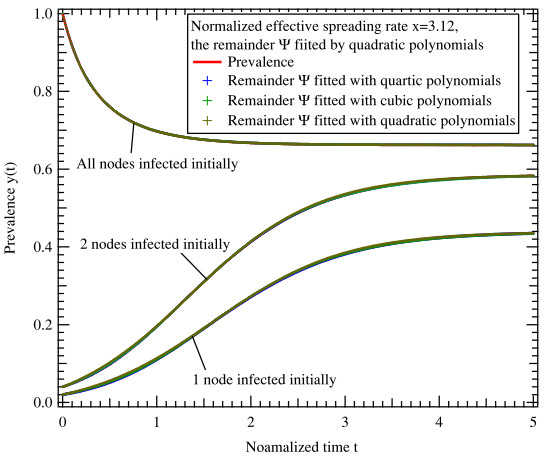
(a) Fitting by quartic polynomials of prevalence y.



(b) Fitting by cubic polynomials of prevalence y.



(c) Fitting by quadratic polynomials of prevalence y.



(d) Solutions of equations in form (5) and the simulated prevalence.

Fig. 8. An example: remainder Ψ corresponding to Fig. 1(a) (above the threshold), fitted by polynomials of prevalence y.

prevalence y. Fig. 9 illustrates the comparison of the prevalence $y(t)$ and the solution of the approximate governing equation $\frac{dy}{dt} = (\tau\mu_{N-1} - 1)y - \tau\mu_{N-1}y^2 - (1.66y^2 - 0.81y - 0.03)$, which corresponds to the fitting curve of the remainder Ψ in Fig. 9. Fig. 9 also demonstrates that the approximate equation fails to describe the epidemic prevalence. Moreover, when the

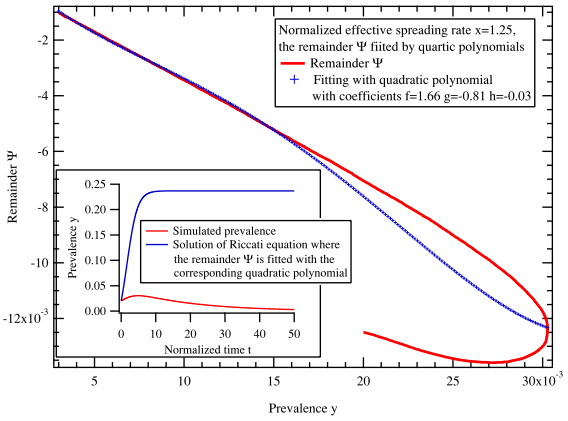
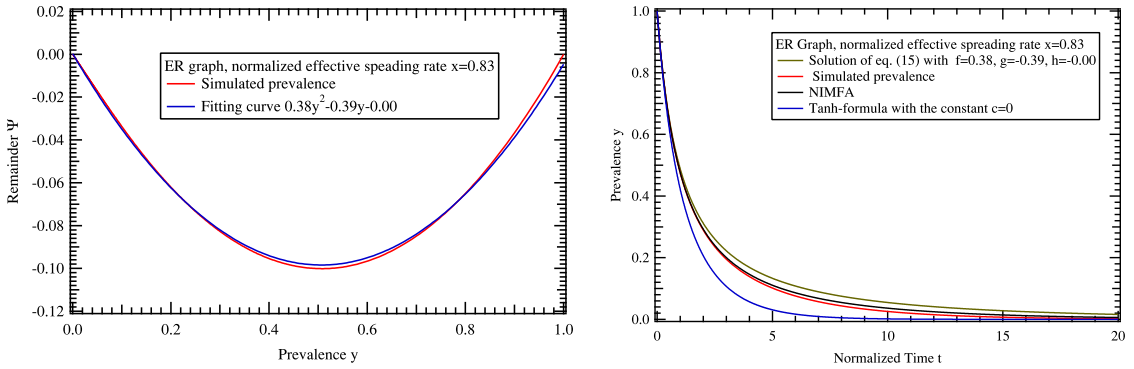


Fig. 9. Remainder Ψ and the quadratic fitting polynomial with normalized effective infection rate $x = 1.25$ (around the threshold) corresponding to Fig. 3(b) of the ER graph.



(a) The remainder Ψ is fitted by quadratic polynomial.

(b) Prevalence comparison.

Fig. 10. Corresponding to Fig. 2 (below the threshold), comparison of the NIMFA prevalence, simulated prevalence and solution approximate equation. The approximate solution shows an improvement in the accuracy than that tanh-formula (3) with the constant $c = 0$.

effective infection rate $\tau \approx \tau_c$, higher order approximate polynomials of the remainder Ψ also cannot improve the accuracy of Eq. (2).

When the effective infection rate $\tau < \tau_c$ and a large number of nodes are infected initially, the remainder $|\Psi|$ can also be large at the initial stage of the epidemic process and $\Psi \not\approx 0$. Fig. 10(a) shows that the remainder Ψ is well fitted with a quadratic polynomial when $\tau < \tau_c$. Furthermore, Fig. 10(b) illustrates the improvement of the approximate governing equation with fitted Ψ, comparing with tanh-formula (3), where the remainder Ψ is considered as a constant c.

We also apply the approximate governing Eq. (6) to the cycle graph, where both the tanh-formula (3) and NIMFA show an inaccuracy in estimating the prevalence. Fig. 11 illustrates the fitting result of the remainder Ψ with the quadratic polynomial $0.15y^2 - 1.00y - 0.03$. As shown in Fig. 11, the solution of the approximate governing equation with fitted Ψ performs better than the tanh-formula (3), where Ψ is considered as a constant c.

5. Conclusion

In this paper, we compare the tanh-formula approximation (3) with the prevalence, which is obtained by averaging 10^6 realizations of the Markovian epidemic process, and the NIMFA prevalence. The results show that, when the effective infection rate τ is above or below the threshold τ_c, the tanh-formula (3) perform well to estimate the time-varying prevalence $y(t)$. If the effective infection rate τ is around the threshold, formula (3) usually fails as well as NIMFA. However, even with the effective infection rate τ is large, the tanh-formula (3) are not accurate in cycle graphs.

The tanh-formula (3) needs an extra fitting parameter c to estimate the time-varying prevalence. The parameter c can be chosen as the remainder Ψ in the metastable state, which may be inaccurate for the prevalence at the initial spreading stage. The remainder Ψ of the Riccati governing Eq. (2) is discussed further. Other than approximating the remainder with a constant $\Psi = c$ in the transient regime, polynomials of the prevalence y are more accurate to approximate the governing Eq. (2) by applying $\Psi \approx f(y)$, where the coefficients of the polynomials are needed as fitting parameters. By fitting the remainder Ψ by a quadratic polynomial, the approximate governing Eq. (6) remains a Riccati differential equation, which is analytically solvable and which describes the Markovian epidemic process better.

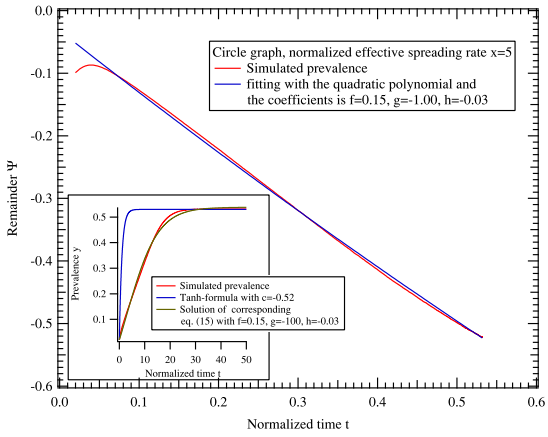


Fig. 11. The remainder Ψ corresponding to Fig. 4(c) with one node infected initially, fitted with quadratic polynomial $0.15y^2 - 1.00y - 0.03$. Also, there is an improvement in accuracy between Eq. (6) with two extra fitting coefficients and the tanh-formula (3) with constant c.

Acknowledgment

Qiang Liu would like to thank the support from China Scholarship Council (No. 201506070094).

Appendix A. Exact Markovian process

In the exact Markovian SIS epidemic model on a graph with N nodes, there are 2^N states, since each node can be either infected or susceptible [17]. Each network state is denoted by an N-digits binary number $x_N(i)x_{N-1}(i)\ldots x_2(i)x_1(i)$, where i is the integer representation of the network state of the Markov process and $i = \sum_{k=1}^{N} x_k(i)2^{k-1}$. The adjacency matrix of the network is $A = \{a_{ij}|1 \le i \le N, 1 \le i \le N\}$.

If a network is in state i at time t, then $X_j(t) = x_j(i)$ for $1 \le j \le N$. The transition between states is a Poisson process [17],

$$\begin{cases} \text{for } j \notin I: & I \to I\bigcup\{j\} \quad \text{at rate } \beta \sum_{k\in I} a_{kj} \\ \text{for } j \in I: & I \to I\backslash\{j\} \quad \text{at rate } \delta \end{cases} \tag{A.1}$$

where I is the set of infected nodes, and $\sum_{k\in I} a_{kj}$ is the number of infected neighbors of node j. The probability state vector $s^T(t) = [s_1(t), s_2(t), \ldots, s_{2^N}(t)]$ represents the probabilities of all possible states of the epidemic in the network at time t. At time t, the probability that the network is in state i is $s_i(t) = \Pr[X_1(t) = x_1(i), X_2(t) = x_2(i), \ldots, X_N(t) = x_N(i)]$. If none of the neighbors of node j is in I, the infection rate $\beta \sum_{k\in I} a_{kj} = 0$ and there is no transition path $I \to I\bigcup\{j\}$. Thus, the state transition graph of a SIS Markov process is determined by the topology of the network.

The exact SIS Markovian process is specified by an $2^N \times 2^N$ linear matrix Q_{SIS} which is the infinitesimal generator. Given the initial state probability vector $s(0)$, the time-dependent state probability vector can be obtained by solving [17],

$$(s'(t))^T = s^T(t)Q_{SIS}. \tag{A.2}$$

From the definition of the joint infection probability $s_i(t)$, by summing over all the $s_i(t)$ with $X_j(t) = 1$, the nodal infection probability of node j can be derived as [10],

$$\Pr[X_j(t) = 1] = E[X_j(t)] = \sum_{i=1}^{2^{N-1}} s_i(t)x_j(i). \tag{A.3}$$

The fraction of infected nodes at time t is $S(t) = \frac{1}{N}\sum_{j=1}^{N} X_j(t)$. The prevalence can be written as $y(t) = E[S(t)] = \frac{1}{N}\sum_{j=1}^{N} E[X_j(t)]$.

The equation set (A.2) contains 2^N linear differential equations, which is infeasible to be calculated to obtain the state probability vector $s(t)$ for realistic network size $N > 15$.

Appendix B. NIMFA prevalence

For a fixed graph, the exact SIS governing equation of node $j(1 \le j \le N)$ is [17],

$$\frac{dE[X_j(t)]}{dt} = -\delta E[X_j(t)] + \beta \sum_{k=1}^{N} a_{kj}E[X_k(t)] - \beta \sum_{k=1}^{N} a_{kj}E[X_j(t)X_k(t)]. \tag{B.1}$$

Eq. (B.1) is not closed due to the existence of joint expectation $E[X_j(t)X_k(t)]$. The governing equation of the joint expectation is derived as [20],

$$\frac{dE[X_iX_j]}{dt} = -2\delta E[X_iX_j] + \beta\sum_{k=1}^{N} a_{ik}E[X_jX_k] + \beta\sum_{k=1}^{N} a_{jk}E[X_iX_k] - \beta\sum_{k=1}^{N}(a_{ik}+a_{jk})E[X_iX_jX_k] \tag{B.2}$$

where 3-variable joint expectations are introduced. Governing equations of nth joint expectations always introduce $(n+1)$th joint expectations. Eventually, the governing equation of the Nth joint expectation is involved, which closes the equations ending up with a total of 2^N equations.

The N-Intertwined Mean-Field Approximation (NIMFA) reduces the computational complexity of the exact Markovian epidemic process (B.1), by assuming independence $E[X_i(t)X_j(t)] = E[X_i(t)]E[X_j(t)]$. Consequently, NIMFA does not calculate joint expectations and the number of equations decreases from 2^N to N. However, the linear Markov equations become non-linear differential equations. The infection probability $v_j(t)$ of NIMFA for node j in G satisfies the differential equation,

$$\frac{dv_j(t)}{dt} = \beta\sum_{k=1}^{N} a_{jk}v_k(t) - v_j(t)\Big(\beta\sum_{k=1}^{N} a_{jk}v_k(t) + \delta\Big) \tag{B.3}$$

which is written in vector form with the NIMFA infection probability vector $V(t) = [v_1(t), v_2(t), \ldots, v_N(t)]$ and the adjacency matrix A, as [17],

$$\frac{dV(t)}{dt} = \beta AV(t) - \mathrm{diag}\big(v_i(t)\big)\Big(\beta AV(t) + \delta u\Big), \tag{B.4}$$

where $\mathrm{diag}\big(v_i(t)\big)$ is a diagonal matrix with elements $v_1(t), v_2(t), \ldots, v_N(t)$ and u is the all-one vector.

The inequality $E[X_i(t)X_j(t)] \geq E[X_i(t)]E[X_j(t)]$ holds for a Markovian SIS epidemic [21], but not necessarily for non-Markovian SIS epidemic spread. Comparing (B.1) and (B.3), we deduce that $\frac{dE[X_i(t)]}{dt} \leq \frac{dv_i(t)}{dt}$ and $E[X_i(t)] \leq v_i(t)$. Hence, the NIMFA prevalence

$$y^{(1)}(t) = \frac{1}{N}\sum_{i=1}^{N} v_i(t) \tag{B.5}$$

upper bounds the exact prevalence $y(t)$.

Appendix C. The SSIS simulator

Due to the infeasible computation of the 2^N states Markov chain, the exact prevalence of a Markovian epidemic process is unavailable. The Simulator of SIS (SSIS), which is a continuous-time Markov process simulator, is applied to obtain the prevalence [22]. The SSIS simulator can simulate the epidemic process with different topologies and parameters. For each run of the simulator (or realization of the SIS process), the fluctuating time-dependent fraction of infected nodes is obtained. Since we can never obtain the exact prevalence $y(t) = E[S(t)]$ without solving approximately 2^N equations for general graphs, here we consider the simulated prevalence as accurate enough after averaging the results of 10^6 realizations of the fractions of infected nodes.

The simulator is based on the Gillespie Algorithm which is used to simulate chemical reactions [23,24]. In the SSIS simulator, the infection and curing events of each node happen randomly on a time line. Since both the infection and curing process are Poisson processes, the length of the time interval between two consecutive events is exponentially distributed. By generating all the consecutive events with an exponentially distributed random time interval on the time line, the network state can be obtained at any arbitrary time. There are two kinds of exponentially distributed random time intervals: the curing time and the infection time. For example, once a node j is infected at time t_0, a random curing time T_C is generated and a curing event is marked on the time line for node j at time $t_0 + T_C$. At time $t_0 + T_C$, node j will be cured and return to the healthy state. Meanwhile, a random infection time T_I is also generated for every neighbor of node j. Furthermore, if the infection time T_I is generated for a specific neighbor k and $T_I < T_C$, an infection event of node k is marked on the time line at time $t_0 + T_I$, which means that, if node k is healthy at $t_0 + T_I$, then node k will be infected. If $T_I > T_C$, no infection event is marked on the time line, because before neighbor k can be infected, node j has already been cured and lost the ability of infection. At time $t = 0$, we choose $Ny(0)$ (the number of initially infected nodes $Ny(0)$ is chosen as an integer in $\{1, 2, \ldots, N\}$) random nodes in the network and mark infection events on them. By generating the events of infection and curing on the time line in chronological order starting from time $t = 0$, the network state $w(t)$ can be obtained at an arbitrary time t.

References

[1] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford, U.K, 1991.
[2] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Modern Phys. 87 (3) (2015) 925–979.
[3] P. Van Mieghem, J. Omic, R. Kooij, Virus spread in networks, IEEE/ACM Trans. Netw. 17 (1) (2009) 1–14.
[4] Y. Wang, D. Chakrabarti, C. Wang, C. Faloutsos, Epidemic spreading in real networks: an eigenvalue viewpoint, in: 22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings, 2003, pp. 25–34.
[5] R. Pastor-Satorras, A. Vespignani, Epidemic dynamics and endemic states in complex networks, Phys. Rev. E 63 (6) (2001) 066117.
[6] P. Van Mieghem, Approximate formula and bounds for the time-varying SIS prevalence in networks, Phys. Rev. E 93 (5) (2016) 052312.
[7] P. Van Mieghem, F. Darabi Sahneh, C. Scoglio, Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold, in: Proceedings of the 53rd IEEE Conference on Decision and Control (CDC14), December, Los Angeles, CA, USA, 2014, pp. 15–17.
[8] P. Van Mieghem, Graph Spectra for Complex Networks, Cambridge University Press, Cambridge, U.K, 2011.
[9] P. Holme, J. Saramki, Temporal networks, Phys. Rep. 519 (3) (2012) 97–125.
[10] P. Van Mieghem, E. Cator, Epidemics in networks with nodal self-infection and the epidemic threshold, Phys. Rev. E 86 (1) (2012) 016116.
[11] E. Cator, P. Van Mieghem, Susceptible-infected-susceptible epidemics on the complete graph and the star graph: exact analysis, Phys. Rev. E 87 (1) (2013) 012811.
[12] Q. Liu, P. Van Mieghem, Die-out Probability in SIS Epidemic Processes on Networks, arXiv:1609.04880.
[13] A.L. Barabsi, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[14] J.M. Hernandez, T. Kleiberg, H. Wang, P. Van Mieghem, 2007, A Qualitative Comparison of Power Law Generators, in: International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), San Diego, California, USA, Jul. 2007, pp. 16–18.
[15] P. Van Mieghem, Decay towards the overall-healthy state in SIS epidemics on networks, arXiv:1310.3980, Oct. 2013.
[16] R. van de Bovenkamp, P. Van Mieghem, Time to metastable state in SIS epidemics on graphs, in: 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), 2014, pp. 347–354.
[17] P. Van Mieghem, Performance Analysis of Complex Networks and Systems, Cambridge University Press, Cambridge, U.K, 2014.
[18] J.R.G. de Mendona, Precise critical exponents for the basic contact process, J. Phys. A: Math. Gen. 32 (44) (1999) L467–L473.
[19] W.G. Cochran, The χ^2 Test of Goodness of Fit, Ann. Math. Statist. 23 (3) (1952) 315–345.
[20] E. Cator, P. Van Mieghem, Second-order mean-field susceptible-infected-susceptible epidemic threshold, Phys. Rev. E 85 (5) (2012) 056111.
[21] E. Cator, P. Van Mieghem, Nodal infection in Markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated, Phys. Rev. E 89 (5) (2014) 052802.
[22] R. van de Bovenkamp, Epidemic processes on complex networks: modelling, simulation and algorithms (Ph.D. Dissertation), Delft University of Technology, The Netherlands, 2015.
[23] D.T. Gillespie, A. Hellander, L.R. Petzold, Perspective: Stochastic algorithms for chemical kinetics, J. Chem. Phys. 138 (17) (2013) 170901.
[24] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361.