Physica A 494 (2018) 317–330

Contents lists available at ScienceDirect

Physica A

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

The spreading time in SIS epidemics on networks

Zhidong He *, Piet Van Mieghem

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

HIGHLIGHTS

- The spreading time resembles a lognormal-like distribution with deep tails.
- The average spreading time is not monotonous with the effective infection rate.
- The average spreading time scales logarithmically as a function of the network size.

ARTICLE INFO

Article history:
Received 14 June 2017
Received in revised form 5 October 2017
Available online 11 December 2017

Keywords:
SIS epidemics
Spreading time
Heavy-tailed distribution

ABSTRACT

In a Susceptible–Infected–Susceptible (SIS) process, we investigate the spreading time T_m, which is the time when the number of infected nodes in the metastable state is first reached, starting from the outbreak of the epidemics. We observe that the spreading time T_m resembles a lognormal-like distribution, though with different deep tails, both for the Markovian and the non-Markovian infection process, which implies that the spreading time can be very long with a relatively high probability. In addition, we show that a stronger virus, with a higher effective infection rate τ or an earlier timing of the infection attempts, does not always lead to a shorter average spreading time $E[T_m]$. We numerically demonstrate that the average spreading time $E[T_m]$ in the complete graph and the star graph scales logarithmically as a function of the network size N for a fixed fraction of infected nodes in the metastable state.

1. Introduction

Epidemic spreading on networks is a ubiquitous process, which can describe the information spreading on social networks [1], emotions [2], biological diseases [3] and failures in networked systems [4]. The Susceptible–Infected–Susceptible (SIS) model is a simple epidemic model where each infected item can be cured, and becomes susceptible again after recovering from the disease. Since the epidemic is a time-dependent spreading process, we are naturally concerned with characteristic times that can be applied to predict or control the spreading process. In spite of the simplicity of the SIS process, unfortunately, only a few results for exact SIS times on a generic graph have been presented [5, p. 460].

In the Susceptible–Infected–Susceptible (SIS) epidemics on a graph, the ratio between the infection rate β and the curing rate δ is called the effective infection rate $\tau = \beta/\delta$. The SIS model features a phase transition [6] around the epidemic threshold τ_c. Viruses with an effective infection rate τ above the epidemic threshold τ_c can infect a sizeable portion of the population on average and stay for a long time in the network. This long period is called the metastable state. Specially, in the Markovian SIS model, the infection processes and the curing processes are Poissonian. A first-order mean-field

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

https://doi.org/10.1016/j.physa.2017.12.048
0378-4371/

approximation of the epidemic threshold $\tau_c^{(1)} = 1/\lambda_1(A)$, where $\lambda_1(A)$ is the spectral radius of the adjacency matrix A, was shown [3,7] to be a lower bound for the epidemic threshold, $\tau_c^{(1)} < \tau_c$.

Due to the existence of an absorbing state, which is the overall healthy or disease-free state in the SIS process, any initial infection will ultimately extinguish in any finite graph. The time until the network reaches the all-healthy state is called the extinction time, or alternatively, the time to absorption or the survival time [8]. When the effective infection rate τ is below the epidemic threshold τ_c, the infectious process dies out exponentially fast [9,10], which is called quick die out or *early* extinction. A sufficient condition for *slow* die out [11] is that the effective infection rate τ is above the epidemic threshold τ_c. If the effective infection rate $\tau > \tau_c$, the infection stays very long on average in any sufficiently large network [12]. The average survival time is dominated by the second largest eigenvalue of the infinitesimal generator of the Markov chain [8,13].

In real-world large graphs, the extinction time is much longer than the actually observed time that an epidemic lasts. Therefore, besides the extinction time, we are interested in characteristic times before the absorbing state is reached. Van de Bovenkamp and Van Mieghem [14] showed that the average hitting time to the metastable state can be computed by using a uniformed embedded Markov chain for the complete graph and the star graph. The modified SIS model in [14] removes the absorbing state directly, implying that the process prevents itself from extinction and restarts to reach the metastable state, from one infected node. Thus, the average time to the metastable state is slightly overestimated, because the restarted process with one infected node usually needs a longer time to reach the metastable state.

In this paper, we define the spreading time T_m as the time when the number I_m of infected nodes in the metastable state is first reached, starting from one initially infected node. The spreading time indicates the spreading velocity of the SIS process in the early stage and unveils the transient, time-dependent properties of epidemic activity before the metastable state. In practice, the average spreading time reflects the time interval in which the virus can be eradicated relatively easily.

Though it is intractable to estimate the spreading time in a general graph analytically, we study the distribution of the spreading time and the factors that influence the spreading time. Based on the simulations, we investigate the distribution of the spreading time T_m both for the Markovian and non-Markovian infection process, and further investigate the effect of the effective infection rate τ, the network size N and the non-Markovian process on the average spreading time $E[T_m]$.

This paper is organized as follows. Section 2 introduces the definition and determination of the spreading time. We investigate the distribution of the spreading time T_m in Section 3. In Section 4, we further present the effect of the effective infection rate τ, the non-Markovian infection times and the network size N on the average spreading time. We conclude the paper in Section 5. We define the metastable state and the stability t_s in a SIS process in Appendix A. Appendix B presents the procedure of the simulator for SIS epidemics (SSIS).

2. Definition and determination of the spreading time

We first propose a preferred definition of the metastable state and the stability time t_s as follow:

Definition 1. In an epidemic process, the metastable state is reached at the stability time t_s, which is the smallest time obeying $\frac{d\overline{y}(t)}{dt}\big|_{t>t_s} < \epsilon$, where the average fraction of infected nodes is $\overline{y}(t) = \frac{1}{N}E[I(t)]$, with $I(t) \geq 1$ is the number of infected nodes at time t, and ϵ is a small positive real number that needs to be agreed upon.

A more detailed discussion on the determination of the stability time is presented in Appendix A.

Definition 2. The spreading time T_m is defined as the first time when the number $I_m = I(t_s)$ of the infected nodes in the metastable state is reached, starting from one initially infected node.

Specifically, the probability distribution of the spreading time T_m in the graph G with N nodes follows

$$\Pr[T_m \leq t] = \sum_{n=1}^{N} \Pr[T_m \leq t | I(t_s) = n] \Pr[I(t_s) = n]. \tag{1}$$

Thus, the average spreading time $E[T_m]$ follows from (1) as

$$E[T_m] = \sum_{n=1}^{N} E[T_{H_n}] \Pr[I(t_s) = n]. \tag{2}$$

where the hitting time $T_{H_n} = T_m\big|_{I(t_s)=n}$ is the first time when the process reaches the state with n infected nodes. After differentiating both sides of (1) with respect to t, we obtain the probability density function (pdf) of the spreading time $f_{T_m}(t)$:

$$f_{T_m}(t) = \sum_{n=1}^{N} f_{T_m}(t | I(t_s) = n) \Pr[I(t_s) = n]. \tag{3}$$

Physically, the spreading time T_m describes the spreading velocity in the early stage of the spreading process, which depends on the local topology around the initial spreaders. After T_m time units, the epidemic approximates the metastable

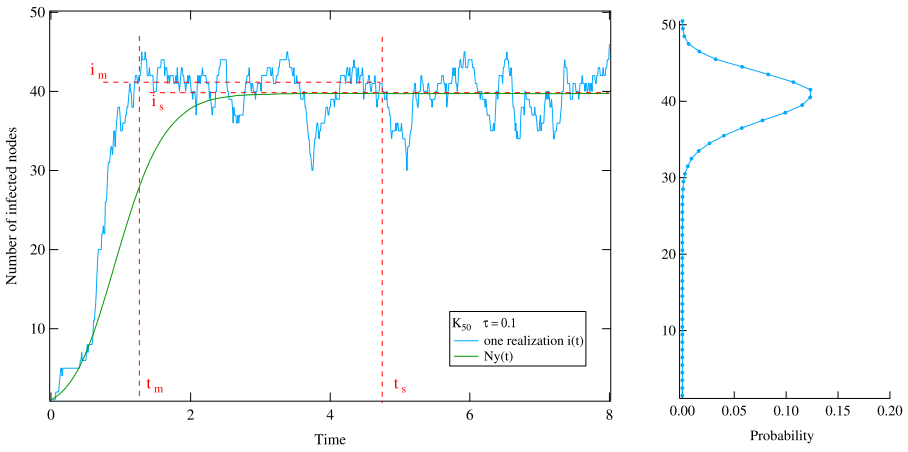

Fig. 1. Illustration of the estimation scheme of the stability time t_s in the prevalence via SSIS and the spreading time t_m for one realization $i(t)$. The distribution of the number of infected nodes in the metastable state is shown in the right subgraph. The green line represents the average number of infected nodes with time based on 10^6 realizations. The time is measured in units of $1/\delta$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

state and already infected a substantial part of the population. Thus, the action of control is preferred to be taken earlier than the average spreading time $E[T_m]$. The average spreading time together with the expected number of infected individuals in the metastable state can guide public health officials in establishing the amount of resources and the available time for the implementation of their mitigation strategies.

Due to the limitation of the analytical methods, an event-driven simulator SSIS (see Appendix B) for the SIS spreading process on a network is implemented based on the Gillespie algorithm [15] to estimate the spreading time. For an unaltered graph and a fixed effective infection rate τ, the epidemic begins with one initially infected node and lasts for the period of t_{limit} time units which is ensured to be long enough to make the spreading process reach the metastable state but not the absorbing state. We record every time point t_k when the kth event happens, as well as the corresponding number of the infected nodes $i(t_k)$ immediately after the kth event. Assume that $0 < t_1 \leq t_2 \leq \cdots \leq t_m < t_{limit}$, then m events have occurred on the timeline before the time limit t_{limit}. After identifying the metastable state and the stability time t_s (see Appendix B), we then determine the spreading time t_m in each realization. The spreading time can be determined from the time t_m when the number of infected nodes $i(t_m)$ first equals to the number $i(t_s)$ of infected nodes at the stability t_s of the metastable state. The random variable T_m corresponds to the spreading time t_m in all realizations that do not go extinct. Fig. 1 illustrates the estimation scheme of the spreading time t_s in a complete graph K_{50}, which also shows the Gaussian-like distribution of the number of infected nodes in the metastable state.

3. Distribution of the spreading time T_m

We first investigate the distribution of the spreading time T_m in the Markovian SIS process. The hitting time T_{H_i} is the first time when the Markov process reaches the state with i infected nodes, starting from one initial spreader. The epidemic process in the complete graph K_N is a birth and death process. Assume that the time is measured in units of $1/\delta$, the average hitting time $E[T_{H_i}]$ from one initial spreader can be analytically derived [14] as

$$E[T_{H_i}] = \sum_{j=1}^{i-1} \sum_{k=0}^{i-j-1} \frac{(N-i+k)!\tau^{j+k-i}}{j(N-j)!} \tag{4}$$

in the modified SIS (MSIS) model [16], where the absorbing state is removed in MSIS Markovian chain. However, a hitting time analysis is tractable when the spreading process can be described as a simple, analytically tractable Markov chain [14].

Fig. 2 exemplifies the average hitting time $E[T_{H_i}]$, from one initial spreader, as a function of the fraction $y = \frac{i}{N}$ of the infected nodes in the complete graph K_{50} with different effective infection rate τ. NIMFA approximates the average number of infected nodes in the metastable state for a complete graph K_N with N nodes as $i_s = \lfloor N\left(1 - \frac{1}{\tau(N-1)}\right)\rfloor$. When the effective infection rate τ is above the epidemic threshold τ_c, the average hitting time $E[T_{H_i}]$ exhibits two different regimes in the average fraction y of infected nodes as shown in Fig. 2. In Regime 1, where $y < \frac{i_s}{N}$, the average hitting time $E[T_{H_i}]$ increases exponentially-like as $e^{\kappa y}$, where the rate κ decreases with the effective infection rate τ. In Regime 2, where $y > \frac{i_s}{N}$, the average hitting time $E[T_{H_i}]$ increases faster than an exponential function.

Fig. 2 suggests that the average hitting time $E[T_{H_n}]$ scales approximately exponentially with the number n of infected nodes around the average number $E[I(t_s)]$ of infected nodes in the metastable state. Assuming that the hitting time T_{H_n}

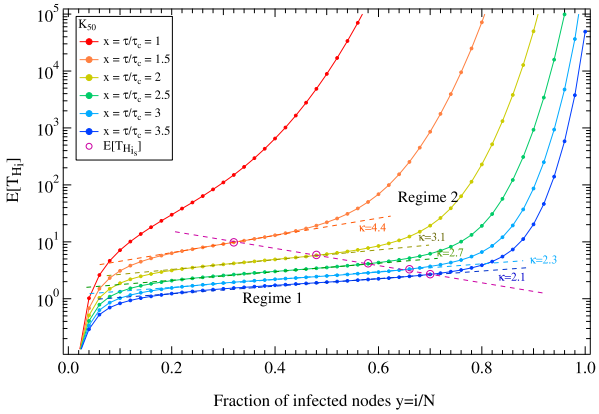

Fig. 2. The average hitting time $E[T_{H_i}]$ to the state with i infected nodes in the complete graph K_{50} with different effective infection rate τ, given that there is one initially infected node. The average fraction of infected nodes in the metastable state via NIMFA is marked.

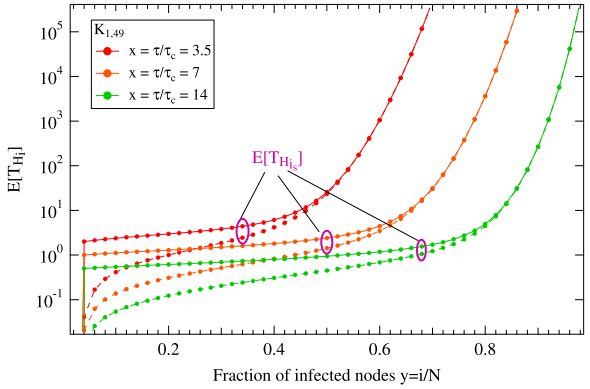

Fig. 3. The average hitting time $E[T_{H_i}]$ to the state with i infected nodes in the star graph $K_{1,49}$ with different effective infection rate τ, given that there is one initially infected node [14]. The solid line represents the process started from a leaf, and the dash line represents the process started from the center. The average fraction of infected nodes $\frac{i_s}{N}$ in the metastable state via NIMFA is marked.

with small variance is correlated to the number n of infected nodes $T_{H_n} \propto e^{\kappa n}$, the spreading time can be regarded as the random variable $T_m(I(t_s)) \approx e^{\kappa I(t_s)+b}$, where the number of the infected nodes $I(t_s)$ is approximately a Gaussian-like random variable [16] with probability density function $\Pr[I(t_s) = n] \approx \frac{1}{\tilde{\sigma}\sqrt{2\pi}} \exp\left[-\frac{(n-\tilde{\mu})^2}{2\tilde{\sigma}^2}\right]$. Therefore, we may infer that the pdf of the spreading time is approximately given by

$$f_{T_m}(t) \approx \frac{1}{\kappa t\tilde{\sigma}\sqrt{2\pi}} \exp\left[-\frac{(\frac{1}{\kappa}(\log t - b) - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right] = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\frac{(\log t-\mu)^2}{2\sigma^2}}, \tag{5}$$

which is a lognormal distribution by replacing $\mu = \kappa\tilde{\mu} + b$ and $\sigma = \kappa\tilde{\sigma}$.

We first show the spreading time T_m started from one initially infected node in two typical graphs including a complete graph K_{50} and a star $K_{1,49}$ with $N = 50$ nodes. Figs. 4 and 5 show the spreading time T_m for two values of normalized effective infection rate $x = \tau/\tau_c$ on a log–log scale, based on more than 10^7 realizations. For both graphs, the distribution of the spreading time is fitted by a lognormal pdf (5) well around the peak probability, with some deviations in the tail. The positive skewness of the distribution, shown in Figs. 4–5, means that the average spreading time $E[T_m]$ is above the mode of the spreading time, which is caused by the rapidly increasing average hitting time $E[T_{H_n}]$ in (2), when the number of infected nodes n exceeds the average number i_s of infected nodes in the metastable state. Comparing the distributions with different normalized effective infection rate x in Figs. 4 and 5, the probability of the small value of the spreading time T_m decreases or even disappears with increasing effective infection rate τ.

Further, Figs. 6–8 show the distributions of the spreading time T_m in an Erdős–Rényi (ER) random graph, a rectangle lattice with $N = 50$ nodes and a BA (Barabási–Albert) power law graph with $N = 1000$ nodes, respectively, where the distribution of the spreading time is influenced by the position of the initially infected spreader and the effective infection rate τ. Taking the lognormal distribution as a reference distribution in the quantile–quantile plots, we find the spreading time also fits the lognormal pdf well when the value of the spreading time is not very large, but deviates in the tail, with a heavier tail

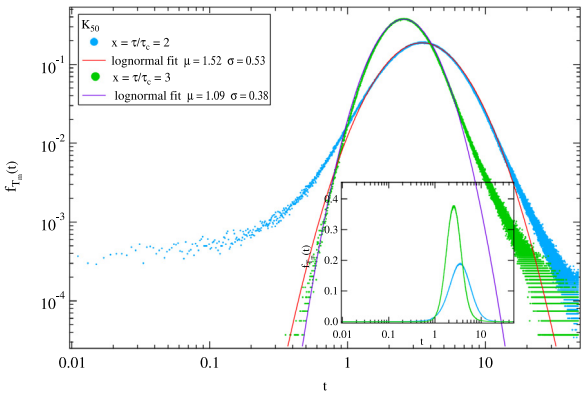

Fig. 4. The distribution of spreading time T_m in the complete graph K_{50} with the effective infection rate x, which is based on more than 5×10^7 realizations. Both axes are on log-scale while only x-axis in the subgraph is on log-scale. The skewness of the distribution is 4.8 for $x = 2$ and 11.4 for $x = 3$.

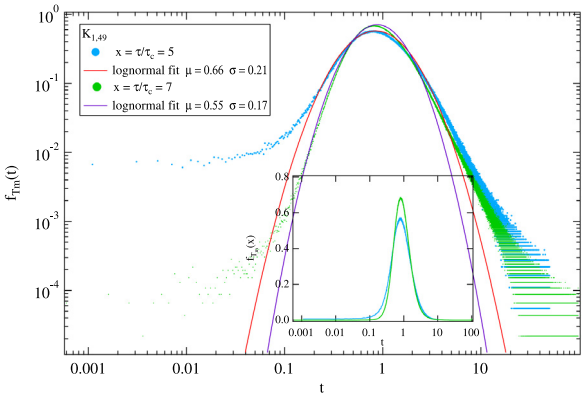

Fig. 5. The distribution of spreading time T_m in the star $K_{1,49}$ with the effective infection rate x, which in based on more than 10^7 realizations. Both axes are on log-scale while only x-axis in the subgraph is on log-scale. The skewness of the distribution is 11.4 for $x = 5$ and 15.3 for $x = 7$.

than the lognormal distribution. Fig. 6 presents the distribution of the spreading time T_m in $10^3 \sim 10^7$ realizations for a connected ER random graph $G_{0.2}(50)$. We observe that the deep tails can be reached only when the number of realizations is extremely large (over 10^6 realizations). If the number of realizations is not large enough, the spreading time is restricted around its average without extreme values. Then, the good fit of the distribution by a lognormal pdf may lead to an *incorrect* conclusion that the spreading time is precisely lognormal.

We also observe that the more regular the graph is, the better the distribution of spreading time T_m fits a lognormal pdf. That regularity agrees with the governing rule of a lognormal, as the limit distribution of a sum of the logarithm of random variable that each does not differ much [5]. In the star or the power-law graph, viruses usually need more time to infect one more node with a very small degree. Fig. 3 for a star graph shows that the function of the hitting time T_{H_i} as the number of infected nodes i increases faster than an exponential around i_s, which may lead to a heavier tail in the distribution of the spreading time, as shown in Fig. 5. We also mark the stability time t_s via simulation in Figs. 6–7, which shows that the stability time t_s lies closely to the tail of the distribution of the spreading time T_m, and is larger than the average spreading time $E[T_m]$.

The infection time is exponentially distributed in the classic Markovian SIS process. More generally, we extend the investigation of the spreading time T_m in a non-Markovian process, which is more common in real-world situations, such as information spread in online social networks and real diseases with incubation periods [17]. We assume that the infection and curing processes are independent in a non-Markovian SIS model, where the curing process is still Poissionian with rate δ, and the infection process at each node infects its neighbors in a time T that is Weibullean, with the pdf

$$f_T(x) = \frac{\alpha}{b}\left(\frac{x}{b}\right)^{\alpha-1} e^{-(x/b)^\alpha}. \tag{6}$$

In order to compare the Weibull with the exponential distribution, we fix the average infection time to $\frac{1}{\beta}$, so that $b = \left(\Gamma\left(1+\frac{1}{\alpha}\right)\beta\right)^{-1}$. Thus, the shape parameter α tunes the power-law start and the tail of the Weibull distributions with the same mean infection time $E[T] = \frac{1}{\beta}$.

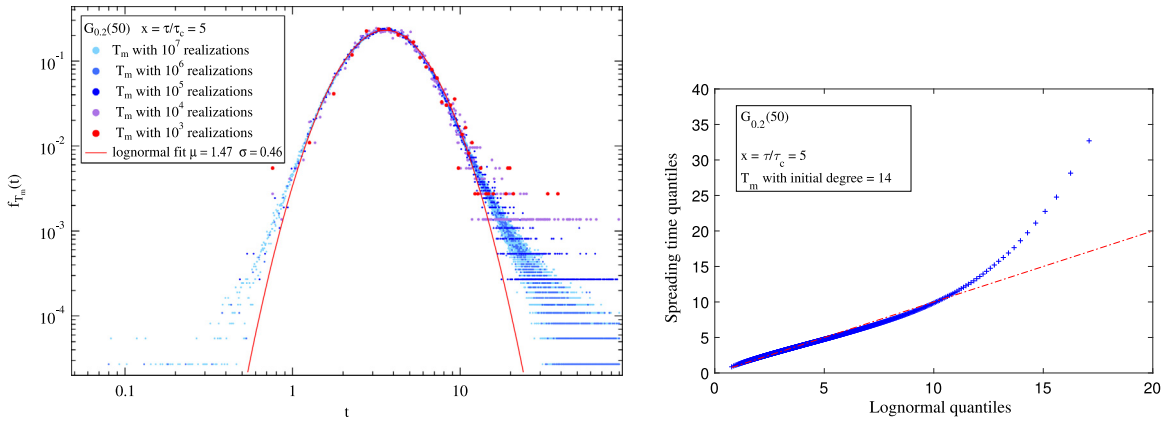

Fig. 6. The distribution of the spreading time T_m in a connected ER random graph $G_{0.2}$ (50) with 50 nodes started from the initial spreader with $d_{initial} = 4$, $x = \tau/\tau_c = 5$.

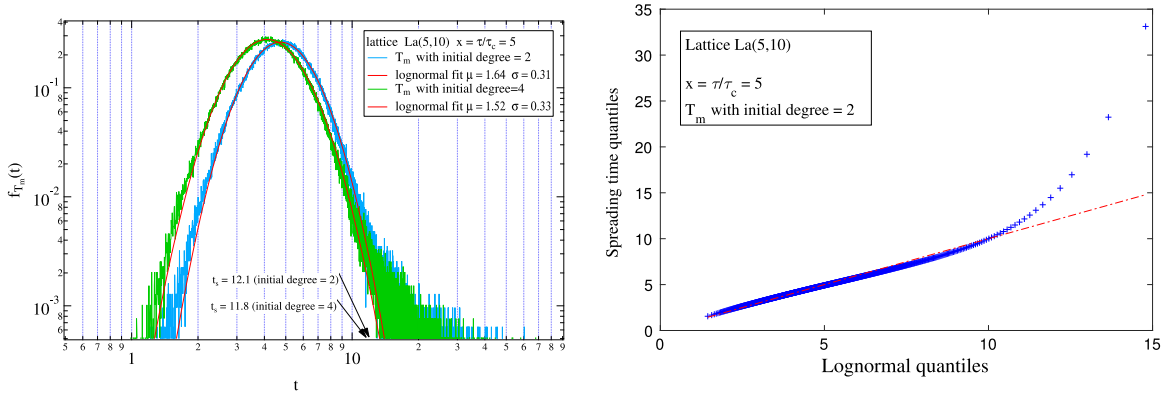

Fig. 7. The distribution of the spreading time T_m in a grid $La(5, 10)$ with 5×10 nodes started from the initial node with $d_{initial} = 2$ and $d_{initial} = 4$, $x = \tau/\tau_c = 5$ where $\tau = 1.35$. The histogram is based on 5×10^5 realizations.

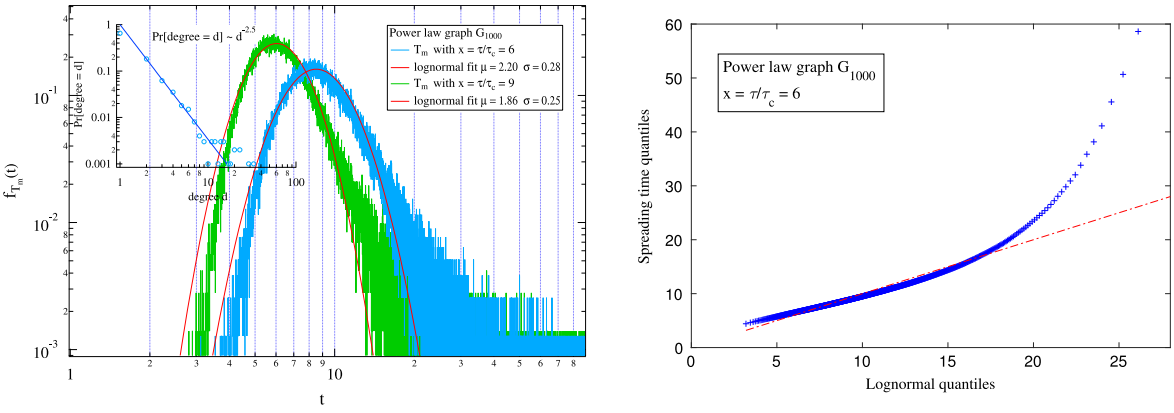

Fig. 8. The distribution of the spreading time T_m in a power law graph G_{1000} with 1000 nodes starting from one initial node. The histogram is based on 2×10^5 realizations.

Figs. 9 and 10 show the distribution of spreading time T_m as a function of the shape parameter α in a complete graph and a star graph. The pdf of the spreading time remains heavy-tailed, and the shape parameter α shifts the mode of the pdf of the spreading time. The tail of the distribution of the spreading time tends to a lognormal pdf better with the increasing shape parameter α in the complete graph.

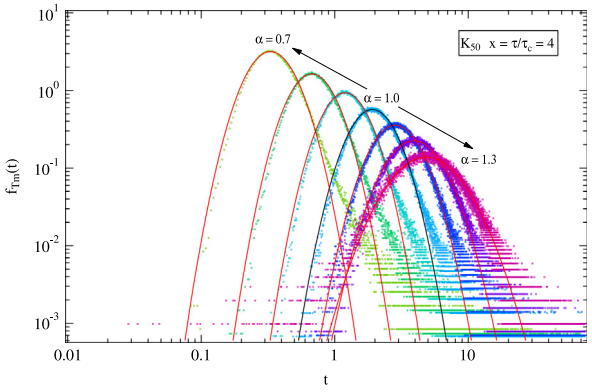

Fig. 9. The distribution of the spreading time T_m with the different shape parameter α in the complete graph K_{50} with the effective infection rate $x = \tau/\tau_c = 4$. The exponential case ($\alpha = 1$) is indicated in black. The histograms are based on more than 5×10^5 realizations.

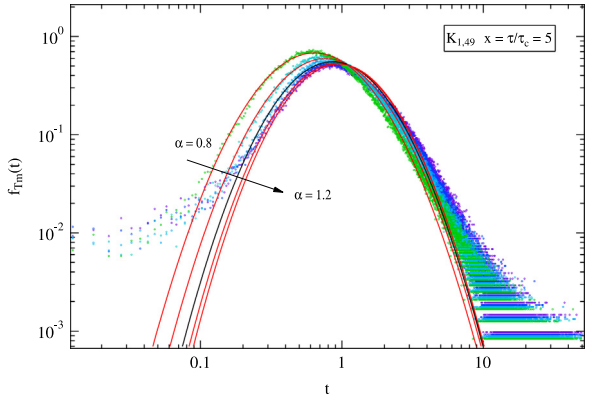

Fig. 10. The distribution of the spreading time T_m with the different shape parameter α in the star graph $K_{1,49}$ with the effective infection rate $x = \tau/\tau_c = 5$. The exponential case ($\alpha = 1$) is indicated in black. The histograms are based on more than 5×10^5 realizations.

The characteristic times with heavy-tailed distribution in Markovian processes have been observed in a few previous research, such as the inter-record time in the extremal process [18], the time of ruin in the risk model [19] and the first return time of random walks [20]. In this section, we show that the spreading time in the SIS model on a network resembles a lognormal-like distribution with different deep tails, regardless of the process being Markovian or non-Markovian, the network topology and the initially infected node.

4. The average spreading time $e[t_m]$ in sis epidemic on networks

4.1. Effect of the effective infection rate on $E[T_m]$

We study the average spreading time $E[T_m]$ as a function of the effective infection rate τ in a SIS process, started from a same initially infected node. Figs. 11 and 12 illustrate the function of the average spreading time $E[T_m]$ with the effective infection rate τ in a complete graph and a star. The average spreading time $E[T_m]$ is not monotonic with the effective infection rate τ but exhibits a maximum, which means that a stronger virus may not lead to a shorter average spreading time $E[T_m]$.

To better explain the above phenomenon, we define the spreading capacity as $c = \frac{E[I_m]}{E[T_m]}$, which approximately indicates the average number of nodes that can be infected in a time unit in the early state of the spreading. Thus, a higher effective infection rate leads to a smaller reciprocal of the spreading capacity $1/c$, which describes the average time units to infect per node. Meanwhile, the average number of infected nodes $E[I_m]$ in the metastable state increases with the effective infection rate τ in a network when the effective infection rate is above the epidemic threshold τ_c. Therefore, the average spreading time $E[T_m]$, which is represented by $E[T_m] = \frac{E[I_m]}{c}$, is influenced by $E[I_m]$ and the spreading capacity c simultaneously, exhibits the property of non-monotony with the effective infection rate τ. The sub-graphs of Figs. 11 and 12 illustrate the reciprocal of the spreading capacity $1/c$ and the average number of infected node $E[I_m]$ in the metastable state as a function of the effective infection rate τ.

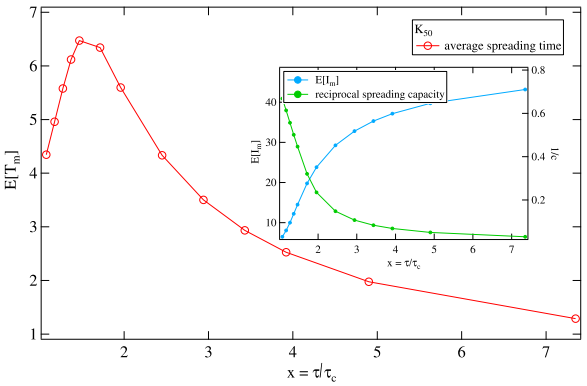

Fig. 11. The average spreading time $E[T_m]$ as a function of the effective infection rate $x = \tau/\tau_c$ in a complete graph K_{50}. The subgraph illustrates the average number $E[I_m]$ of infected nodes in the metastable state and the reciprocal of the spreading capacity $1/c$ with the normalized effective infection rate $x = \tau/\tau_c$.

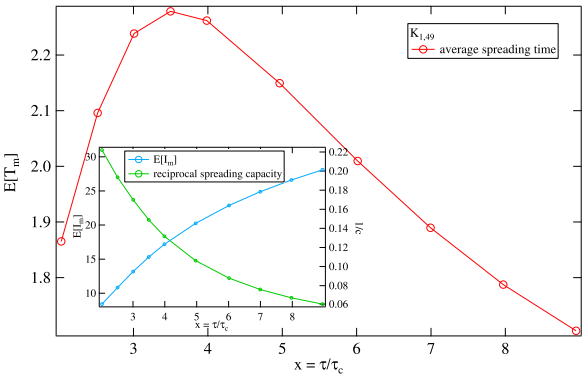

Fig. 12. The average spreading time $E[T_m]$ with $x = \tau/\tau_c$ in a star graph $K_{1,49}$ with 49 leaves, started from the center of the graph. The subgraph illustrates the average number $E[I_m]$ of infected nodes in the metastable state and the reciprocal of the spreading capacity $1/c$ with the normalized effective infection rate $x = \tau/\tau_c$.

4.2. *Effect of the shape parameter α on $E[T_m]$*

We now investigate the effect of the shape parameter α in the Weibull-distributed infection time with pdf (6) on the average spreading time $E[T_m]$, where the Markovian infection process is a special case with $\alpha = 1$. As discussed in Section 4.1, the average spreading time depends on the spreading capacity c and the average fraction $y(t_s)$ of infected nodes in the metastable state, both of which are influenced by the shape parameter α.

The average number of infection attempts during a recovery time is a physically more general description than the effective infection rate in non-Markovian epidemics [17]. Considering the distribution of the infection attempts over an infectious period of a node, the occurrence of events is not uniformly distributed over an interval when the infection process is non-Markovian. For $\alpha < 1$, the infection events tend to happen earlier than the Poisson-distributed events (for $\alpha = 1$) with high probability, while for $\alpha > 1$, the infection events tend to happen later. Therefore, the timing of the infection attempts relative to the curing time of a node influences the epidemics process even for the same average number of expected infection attempts [8]. Physically, the reciprocal of the spreading capacity $1/c$, which describes the average time units to infect per node before the metastable state, also increases for a higher α.

Fig. 13 shows that the average fraction $y(t_s)$ of infected nodes in the metastable state depends on both the effective infection rate τ and the shape parameter α. Specifically, the average fraction $y(t_s)$ of infected nodes in the metastable state decreases with a higher parameter α for a same effective infection rate τ. Fig. 14 suggests that $\log(\tau) \sim \frac{\log(Ny(t_s))}{\alpha}$ for the same number $Ny(t_s)$ of infected nodes in the metastable state, which implies that $\tau^\alpha \sim y(t_s)$ in the complete graph when $\tau < 1$. This relation is consistent with the conclusion that the epidemic threshold $\tau_c(\alpha)$ in the non-Markovian SIS epidemics scales as $(\tau_c^{(1)})^{\frac{1}{\alpha}}$, where $\tau_c^{(1)} = \tau_c(1)$ is the epidemic threshold in the Markovian SIS model [7]. As Fig. 14 shows in the star graph, the Weibull shape factor α barely influences the fraction $y(t_s)$ of infected nodes in the metastable state when the effective infection rate $\tau \geq 1$.

Fig. 15 shows that, both in the complete graph and the star, the average spreading time $E[T_m]$ does not always increase monotonically with the shape parameter α, but exhibits a maximum when the effective infection rate τ is small. For a higher

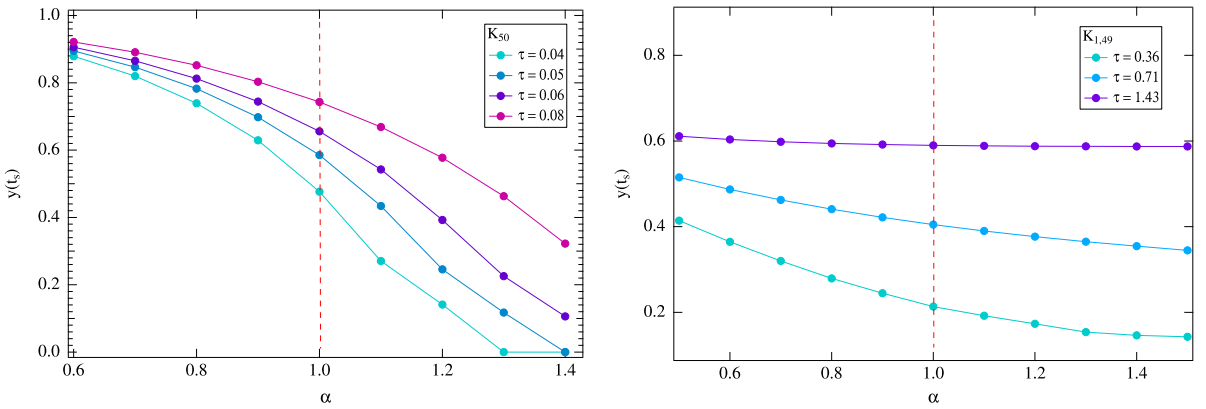

Fig. 13. The average fraction of infected nodes in the metastable state for the same τ in the non-Markovian SIS process in a complete graph K_{50} and a star graph $K_{1,49}$.

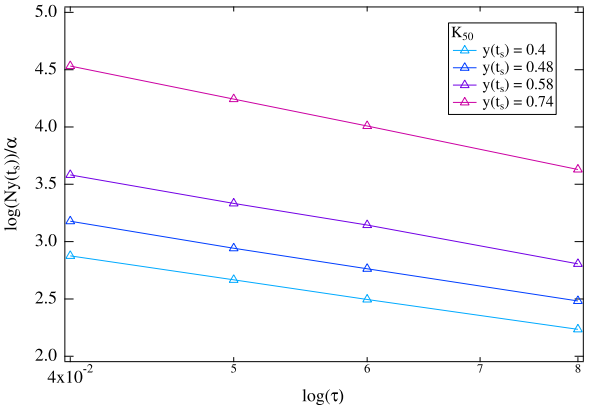

Fig. 14. The reciprocal of the parameter α as a function of $log(\tau)$ in the complete graph K_{50} for the same fraction of infected nodes in the metastable state.

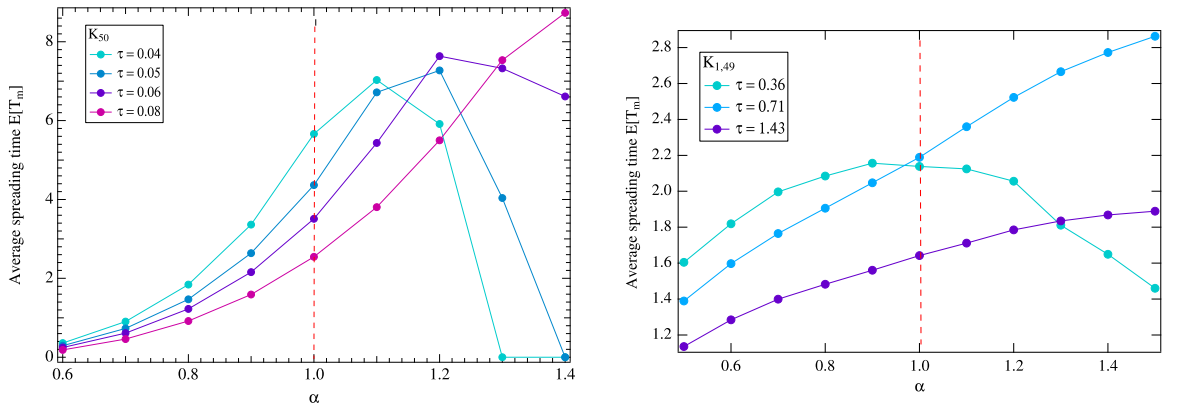

Fig. 15. The average spreading time $E[T_m]$ as a function of the parameter α for the same effective infection rate τ in the complete graph K_{50} and in the star graph $K_{1,49}$.

α, the timing of the infection attempts is postponed while the fraction of infected nodes in the metastable state decreases. These two factors leads to the non-monotonicity of the average spreading time $E[T_m]$ with the shape parameter α, and implies that increasing the parameter α may not shorten the average spreading time $E[T_m]$.

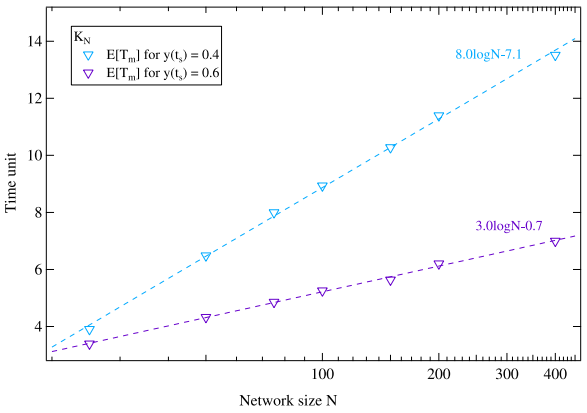

Fig. 16. The average spreading time $E[T_m]$ starting from one initially infected node as a function of the network size in the complete graph K_N.

4.3. Effect of the network size on $E[T_m]$

We now investigate the effect of the network size N on the average spreading time $E[T_m]$. Figs. 16–18 show the average spreading time $E[T_m]$ starting from one initially infected node as a function of the network size for a complete graph K_N, a star $K_{1,N}$, and an ER random graph $G_p(N)$. Referring to the average fraction of infected nodes $y^{(1)}(t_s) = 1 - \frac{1}{(N-1)\tau}$ in the metastable state in a complete graph with N nodes via NIMFA [16], we can estimate the effective infection rate $\tau = \frac{1}{(1-y(t_s))(N-1)}$ for a fixed average fraction $y(t_s)$ of infected nodes in the metastable state. Similarly, in an ER graph, the effective infection rate $\tau = \frac{1}{(1-y(t_s))(N-1)p}$ for a fixed fraction $y(t_s)$ of infected nodes in the metastable state is estimated by the NIMFA approximation $y^{(1)}(t_s) = 1 - \frac{1}{(N-1)p\tau}$, where the link probability $p = \frac{2\log N}{N}$. The effective infection rate τ in a star is estimated by the NIMFA approximation [16] that $y^{(1)}(t_s) = \frac{N-\tau^{-2}}{N+1}\left\{\frac{1}{\tau^{-1}+1} + \frac{1}{\tau^{-1}+N}\right\} \approx \frac{\tau}{1+\tau}$ when $N \gg \tau$.

We ignore the curing events and consider a Susceptible–Infected (SI) process in the complete graph. The average time when I_m nodes are infected [5] follows $\sum_{n=1}^{I_m} \frac{1}{\tau n(N-n)}$, where $y(t_s) = \frac{I_m}{N} \approx 1 - \frac{1}{N\tau}$ is fixed. Thus, we obtain

$$E[T_m] \approx \sum_{n=1}^{I_m} \frac{1}{\tau n(N-n)} = \frac{2}{\tau N}\sum_{n=1}^{I_m}\frac{1}{n} \sim 2(1-y(t_s))\log(y(t_s)N), \tag{7}$$

which scales logarithmically with the network size N. For an SIS process, Figs. 16–18 show that the average spreading time $E[T_m]$ via simulation approximately scales logarithmically as $a\log(N) + b$ for different fractions $y(t_s)$ of infected nodes in the metastable state in a complete graph, an ER random graph and a star. The process needs more time to infect a same fraction of nodes in a network with a larger size. The slope a of the fit is larger for a smaller fraction $y(t_s)$ of infected nodes in the metastable state, which means the average spreading time $E[T_m]$ tends to increase more quickly with the network size N when the fraction $y(t_s)$ of infected nodes in the metastable state is smaller. We observe the similar trend of the average spreading time with the network size N in the complete graph and the ER random graph. Actually, when the link density p in an ER random graph is above the critical link density $p_c = \frac{\log N}{N}$, the graph is already dense and follows similar behaviors as the complete graph [5].

5. Conclusion

We define the spreading time as the time when the number of infected nodes in the metastable state is first reached, starting from the outbreak of an epidemic.

We investigated the distribution of the spreading time. The average hitting time $E[T_{H_i}]$ to the state i around the average number of infected nodes in the metastable state approximates an exponential function, where the number of infected nodes in the metastable state resembles a Gaussian-like distribution. Thus, we observe that the spreading time T_m resembles a lognormal-like distribution with different deep tails, which is exhibited both in the Markovian and the non-Markovian infection process.

We further investigated the properties of the average spreading time. Because the number of infected nodes in the metastable state and the spreading capacity are influenced by the effective infection rate simultaneously, the average spreading time $E[T_m]$ is not necessarily monotonous with the effective infection rate τ but exhibits a maximum, which means that a higher effective infection rate τ may not lead to a shorter average spreading time $E[T_m]$. Similarly, both the fraction of infected nodes in the metastable state and the timing of the infection attempts are influenced simultaneously by the parameter α of the Weibullean infection times, which leads to non-monotonicity of the average spreading time $E[T_m]$

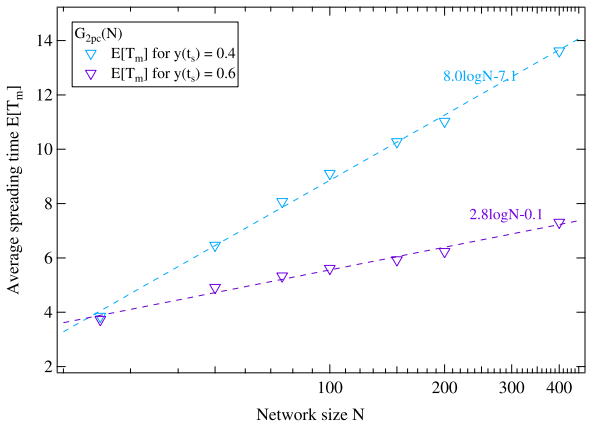

Fig. 17. The average spreading time $E[T_m]$ starting from one initially infected node with the average degree as a function of the network size in the ER random graph $G_{2p_c}(N)$.

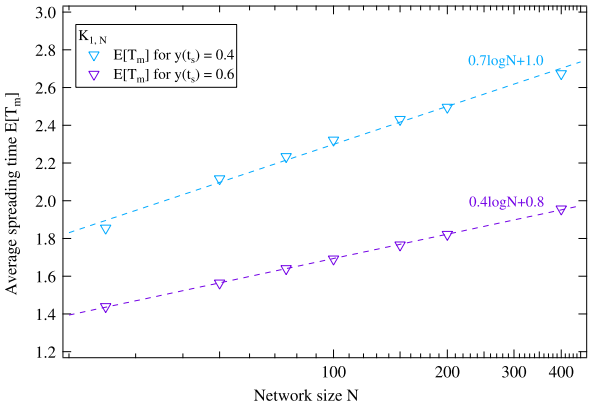

Fig. 18. The average spreading time $E[T_m]$ starting from the center as a function of the network size in the star graph $K_{1,N}$.

with the shape parameter α. Finally, we showed that the average spreading time $E[T_m]$ scales logarithmically as a function of the network size N, given that the average fraction $y(t_s)$ of infected nodes in the metastable state is fixed.

Acknowledgments

This research is supported by the China Scholarship Council (CSC). We are grateful to M. Märtens, H. Cetinay and X. Wang for useful comments.

Appendix A. Determination of the metastable state and the stability time

We define a Bernoulli random variable $X_i(t) \in \{0, 1\}$ as the infectious state of node i, where $X_i(t) = 1$ indicates that node i is infected and $X_i(t) = 0$ indicates that node i is susceptible at time t. The prevalence $y(t) = \frac{1}{N}E[I(t)]$ of an SIS process is the expected fraction of infected nodes at time t, where $I(t) = \sum_{i=1}^{N} X_i(t)$ is the number of infected nodes. We present several definitions of the metastable state in the SIS process on finite graphs derived from the prevalence $y(t)$ in this section.

Definition 1(a). In an epidemic process, the metastable state is reached at the stability time t_s, which is the smallest time obeying $\frac{dy(t)}{dt}\Big|_{t=t_s} = 0$.

It seems reasonable to define the start of the metastable state when the prevalence $y(t)$ reaches its first extremum. However, the SIS prevalence $y(t)$, started from multiple initial spreaders, may pass multiple extrema in the transient regime in a specific network, which demonstrates that Definition 1(a) is not precise. In addition, as shown in Fig. A.19, the prevalence $y(t)$ may monotonically decreases when the average number of infected nodes in the metastable state is smaller than the number of the initially infected nodes. Therefore, this definition may not be adequate for the computation of the spreading, starting from multiple initially infected nodes.

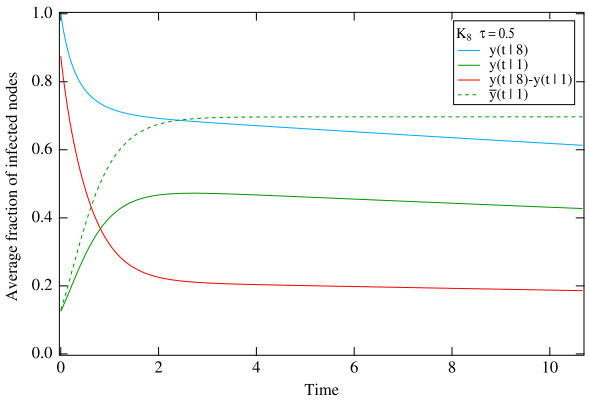

Fig. A.19. The exact prevalences $y(t|1)$ started from one infected node and the exact prevalence $y(t|8)$ started from all infected nodes in a complete graph K_8 with the effective infection rate $\tau = 0.5$. The red line represents the difference between the two prevalences $y(t|1)$ and $y(t|8)$. The green dash line represents the prevalence $\bar{y}(t|1)$ excluding early extinction probability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Definition 1(b). In an epidemic process, the metastable state is reached at the stability time t_s, which is the smallest time obeying $\frac{dy(t)}{dt}\big|_{t=t_s} = 0$, and $|y(t) - y(t_s)| \leq \epsilon$ for $\forall t > t_s + \alpha E[T_{absorbing}]$, where $0 < \alpha < 1$. The positive real numbers α and ϵ need to be agreed upon.

To remedy the defect of Definition 1(a), we try to bound the prevalence $y(t)$ in an interval around the fraction $y(t_s)$ of infected node at the stability time t_s. However, the prevalence $y(t)$ will inevitably exceed the bound because the prevalence will reach an absorbing state $y(t) = 0$ finally. Therefore, it is hard to determine the two parameters ϵ and α that allows $|y(t) - y(t_s)| \leq \epsilon$ for $\forall t > t_s + \alpha E[T_{absorbing}]$.

Definition 1(c). In an epidemic process, the metastable state is reached at the stability time t_s, which is the smallest time obeying $y(t_s|i) = y(t_s|N)$.

Because we cannot bound the prevalence $y(t)$ in the metastable state, we consider to use the prevalence $y(t|N)$ started from $I_0 = N$ initial nodes as a reference curve and locate the start of the metastable state as the intersection point in time with the prevalence $y(t|I_0 = i)$, where the prevalence $y(t|N)$ with all initial spreaders converges fastest to the metastable state.

Fig. A.19 shows that there exists a gap between the prevalences $y(t|1)$ and $y(t|N)$ in the metastable state due to the different probability of extinction, which means that the prevalence started from a different number of initial nodes will not intersect before the absorbing state. Fig. A.19 also shows that the difference between the two prevalences $y(t|1)$ and $y(t|N)$ becomes narrower with the time, which implies that the decreasing rate of the prevalence is also influenced by the initial infection condition. We expect that all the prevalences $y(t|I_0)$ with $I_0 \in (1, 2, \ldots, N)$ initially infected nodes will meet only in the absorbing state, which demonstrates the infeasibility to locate the metastable state by the intersection of the prevalence curves.

Definition 1(d). In an epidemic process, the metastable state is reached at the stability time t_s, which is the first time obeying $\frac{dy(t)}{dt}\big|_{t>t_s} < 0$.

Definition 1(d) means that the last extremum of the prevalence $y(t)$ is located as the start of the metastable state, and the average fraction of infected nodes monotonically decreases after the stability time t_s. The prevalence $y(t)$ is the average fraction of infected nodes, which includes the realizations that die out early as well as the realizations that reach the metastable state, thus

$$y(t) = y(t)\big|_{I(t)>0} \Pr[I(t) > 0] + y(t)\big|_{I(t)=0} \Pr[I(t) = 0]. \tag{A.1}$$

The fraction $y(t)\big|_{I(t)>0}$ approximates gradually the fraction $y(t)$ with the decreasing extinction probability $\Pr[I(t) = 0]$. However, the extinction probability $\Pr[I(t) = 0]$ is hard to estimate in a general network mathematically.

Definition 1(e). In an epidemic process, the metastable state is reached at the stability time t_s, which is the smallest time obeying $\frac{d\bar{y}(t)}{dt}\big|_{t>t_s} < \epsilon$, where the average fraction of infected nodes is $\bar{y}(t) = \frac{1}{N}E[I(t)]$, with $I(t) \geq 1$ is the number of infected nodes at time t, and ϵ is a small positive real number that needs to be agreed upon.

Algorithm 1 Simulation for SIS epidemics

Inputs:
G_N: the network with N nodes;
I_0: the initial spreader(s);
β: the infection rate; δ: the curing rate;
t_{limit}: the time limit; $t_{current}$: the current time;
Outputs:
$i(t)$: the number of infected nodes at time t;
Initialization:
$t_{current} \leftarrow 0$;
Insert the events $\Omega_{n,1}(0)$ for the initially infected nodes on the timeline;
while $t_{current} < t_{limit}$ **do**
  Find the earliest un-handled event $\Omega_n(t)$ on the timeline;
  $t_{current} \leftarrow t$;
  if $\Omega_n(t)$ is an infection event **then**
    if Node n is susceptible **then**
      Node n becomes infected;
      $i(t) \leftarrow i(t) + 1$;
      Insert the event $\Omega_{n,0}(t')$, where $t' \leftarrow t + rand(1/\delta)$ and $rand(1/\delta)$ is an exponentially distributed random time interval with mean $1/\delta$;
      for each neighbor m of node n **do**
        Generate $t'' \leftarrow t + rand(1/\beta)$, where $rand(1/\beta)$ is an exponentially distributed random time interval with mean $1/\beta$;
        if $t'' < t'$ **then**
          Insert the event $\Omega_{m,1}(t'')$;
        end if
      end for
    end if
  else if $\Omega_n(t)$ is a curring event **then**
    Node n is cured;
    $i(t) \leftarrow i(t) - 1$;
  end if
end while

In the above definition, we introduce the prevalence

$$\overline{y}(t) = y(t)\big|_{I(t)>0} = \frac{y(t)}{1 - \Pr[I(t) = 0]} \tag{A.2}$$

subject to the condition that the process does not die out, where $\Pr[I(t) = 0]$ is the extinction probability. The prevalence $\overline{y}(t)$ excluding early extinction, as illustrated in Fig. A.19, tends to stay almost constant instead of decaying as the prevalence $y(t)$ after reaching the extremum. We consider that the metastable state starts when the prevalence $\overline{y}(t)$ stays almost constant. Actually, the prevalence $\overline{y}(t)$ excluding early extinction is a monotonically increasing function, which only stays constant when $t \to \infty$, as follows from general Markov theory [5]. The prescribed stringent parameter ϵ can be determined as a small value. Definition 1(e) is also consistent with the definition of the quasi-stationary state, which leads to the almost steady average number of infected nodes without extinction realizations.

In summary, we choose Definition 1(e) as our preferred definition of the metastable state and the stability time t_s in this paper.

Appendix B. Simulation for a SIS process on networks

There are two kinds of events in the SIS process which are infection events and curing events. All the events are marked on a same timeline and are handled by the order of their time after the beginning of the simulation. We denote by $\Omega_{n,1}(t)$ the infection event that node n becomes infected at time t, and $\Omega_{n,0}(t)$ the curing event that node n becomes cured at time t. The process of SSIS (Simulation for SIS epidemics) is described by Algorithm 1.

We set the parameter $\epsilon = 0.01$ in Definition 1(e) and run the SSIS repeat for an unaltered graph, a fixed effective infection rate τ and the same initial condition. The prevalence can be obtained by $\overline{y}(t) = \frac{1}{N}E[I(t)]$, where the random variable $I(t)$ denotes the number of infected nodes $i(t)$ in all realizations. Then we determine the stability time t_s as the first time when $\frac{\overline{y}(t_s+\Delta t)-\overline{y}(t_s)}{\Delta t} < \epsilon$, where the time sample interval $\Delta t = 0.01$ in our simulations.

References

[1] R. Pastor Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Modern Phys. 87 (3) (2015) 925.
[2] A.L. Hill, D.G. Rand, M.A. Nowak, N.A. Christakis, Emotions as infectious diseases in a large social network: the SISa model, Proc. R. Soc. B 277 (1701) (2010) 3827–3835.
[3] P. Van Mieghem, J. Omic, R. Kooij, Virus spread in networks, IEEE/ACM Trans. Netw. 17 (1) (2009) 1–14.
[4] L. Buzna, K. Peters, D. Helbing, Modelling the dynamics of disaster spreading in networks, Physica A 363 (1) (2006) 132–140.
[5] P. Van Mieghem, Performance Analysis of Complex Networks and Systems, Cambridge University Press, 2014.
[6] C. Castellano, R. Pastor-Satorras, Thresholds for epidemic spreading in networks, Phys. Rev. Lett. 105 (21) (2010) 218701.
[7] P. Van Mieghem, R. van de Bovenkamp, Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks, Phys. Rev. Lett. 110 (10) (2013) 108701.
[8] R. van de Bovenkamp, P. Van Mieghem, Survival time of the susceptible-infected-susceptible infection process on a graph, Phys. Rev. E 92 (3) (2015) 032806.
[9] A. Ganesh, L. Massoulié, D. Towsley, The effect of network topology on the spread of epidemics, in: 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2005, in: Proceedings IEEE, vol. 2, IEEE, 2005, pp. 1455–1466.
[10] P. Van Mieghem, Approximate formula and bounds for the time-varying susceptible-infected-susceptible prevalence in networks, Phys. Rev. E 93 (5) (2016) 052312.
[11] M. Draief, Epidemic processes on complex networks, Physica A 363 (1) (2006) 120–131.
[12] P. Van Mieghem, Decay towards the overall-healthy state in SIS epidemics on networks, 2013. arXiv preprint arXiv:1310.3980.
[13] J.R. Artalejo, On the time to extinction from quasi-stationarity: A unified approach, Physica A 391 (19) (2012) 4483–4486.
[14] R. van de Bovenkamp, P. Van Mieghem, Time to metastable state in SIS epidemics on graphs, in: Tenth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), IEEE, 2014, pp. 347–354.
[15] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361.
[16] 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.
[17] E. Cator, R. van de Bovenkamp, P. Van Mieghem, Susceptible-infected-susceptible epidemics on networks with general infection and cure times, Phys. Rev. E 87 (6) (2013) 062816.
[18] V. Pisarenko, M. Rodkin, Heavy-Tailed Distributions in Disaster Analysis, Springer Science & Business Media, 2010.
[19] D.C. Dickson, H.R. Waters, The distribution of the time to ruin in the classical risk model, Astin Bull. 32 (02) (2002) 299–313.
[20] S. Kostinski, A. Amir, An elementary derivation of first and last return times of 1D random walks, Amer. J. Phys. 84 (1) (2016) 57–60.