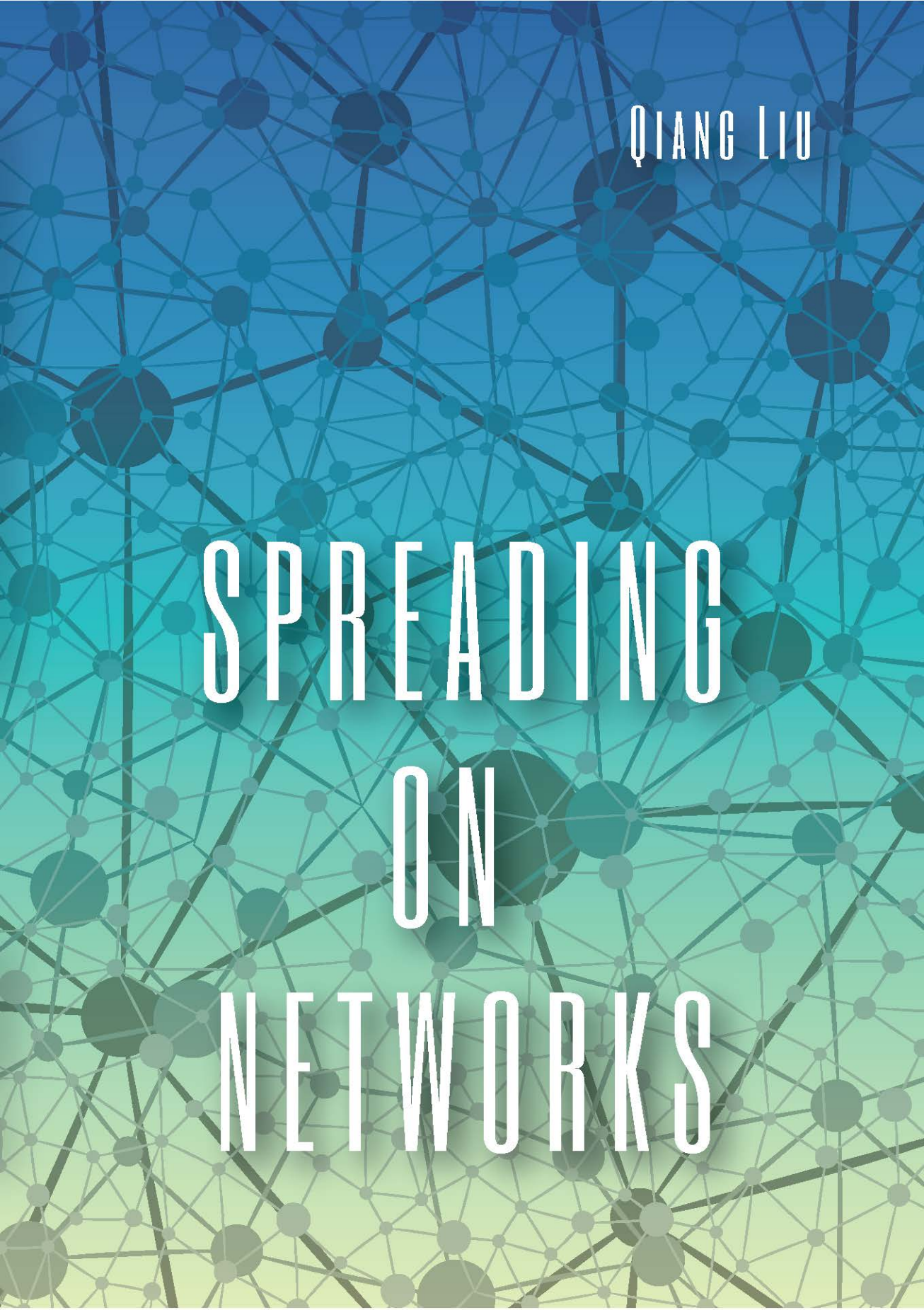
Qiang Liu
SPREADING
ON
NETWORKS

Spreading on Networks

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Thursday 31 October at 10:00 o'clock

by

Qiang LIU

Master of Science in Cryptography,
University of Electronic Science and Technology of China, Chengdu, China,
born in Shandong, China.

This dissertation has been approved by the promotor:
Prof.dr. P. F. A. Van Mieghem

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. P. F. A. Van Mieghem	Delft University of Technology, promotor

Independent members:

Prof. dr. C. M. Scoglio	Kansas State University, United States
Prof. dr. Y. Moreno	University of Zaragoza, Spain
Prof. dr. M. Boguñá Espinal	University of Barcelona, Spain
Prof. dr. R. E. Kooij	Delft University of Technology
Prof. dr. A. Hanjalic	Delft University of Technology
Dr. J. L. A. Dubbeldam	Delft University of Technology

Keywords: Complex Networks, Spreading Process, Stochastic Simulation

Front & Back: Designed by Liangyi Li

Printed by: Ipskamp Printing, Enschede

Published by: Qiang Liu

Email: Q.L.Liu@hotmail.com

URL: http://QiangLiu.net

ISBN 978-94-6384-074-3

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

To my family

Table of Contents

Summary

Spreading phenomena such as spreading of diseases, information and computer viruses are ubiquitous in nature and man-made systems, but the understanding of them is still insufficient. This dissertation focuses on the analysis of a basic mathematical model of spreading phenomena running on underlying network structures and aims to complete the basic theory of spreading processes. Specifically, we explore the Susceptible-Infected-Susceptible (SIS) model from several interesting perspectives to contribute to the state-of-the-art understanding of the model.

Our first main contribution is related to temporal correlations. In most of the studies, the influence of time in the SIS spreading process is omitted because the specific value of the infection and curing rates does not influence the first-moment metastable properties, such as the infection probability of each node. Only the ratio between the two rates matters. In this dissertation, we show that the temporal correlation can be analyzed with the mean-field approaches, although mean-field methods are meant to only analyze first-moment properties. We derive the autocorrelation of the nodal infection state both in the steady and transient states under the mean-field approximation. By analyzing the autocorrelation, we indicate the influence of the underlying network and the value of the infection and curing rates on the temporal properties of the spreading process. We also show that the infection and curing rates can be calculated by measuring the infection state of each node.

Second, we relax the Markovian assumption in the SIS process by extending the Poisson infection process to a Weibull renewal process. The Poisson infection process is just a special case of the Weibullian renewal process. Under this Weibullian framework, we can parameterize the non-Markovian infection behavior and show some new features raised by it. We specifically focus on an extreme (limiting) case of the Weibullian SIS process where the distribution of the infection time is a Dirac delta function. The analysis of the extreme case leads to the largest possible epidemic threshold for non-Poissonian infection processes. We further discuss the epidemic threshold for different infection processes with Weibull, lognormal and Gamma distributed infection time, which fit realistic spreading phenomena well, under a previous non-Markovian mean-field method based on renewal theory. We show consistency between our results and previous theory and that those different infection processes behave similarly.

Third, we dive into the localization phenomena in networks from the view-

point of SIS spreading processes. Localization of the spreading process appears just above the epidemic threshold in networks whose principal eigenvector of the adjacency matrix is localized. In the localized spreading, the prevalence (order parameter), which is the expected fraction of infected nodes, converges to zero with the increase of network size but the number of infected nodes is non-zero. Thus, the localized spreading forms an interesting phase different from the all-healthy phase (no infection) and the endemic phase (non-zero prevalence). We evaluate the above-mentioned extreme case of the Weibullian SIS process where the time-dependent prevalence is periodic in the long-run. Near the epidemic threshold, the ratio between the steady-state maximum and minimum prevalence, which equals to the largest eigenvalue of the adjacency matrix, diverges in some networks, but the spreading process is still localized. In other words, the divergent ratio of prevalence, determined by the largest eigenvalue of the network, cannot amplify a zero-prevalence to a non-zero one in the thermodynamic limit. The result indicates that the localization of spreading processes may be only determined by the network structure but not the specific infection process.

Finally, we study the curing strategy for the control of the spreading process, specifically, the pulse curing strategy. Compared to the classical asynchronous curing strategy (for instance Poissonian), pulse strategy is an optimized method of suppressing the spreading and applied broadly in disease control. Here, we study the model which is composed of a susceptible-infected process and a periodical pulse curing process with a successful curing probability below one. We derive the mean-field epidemic threshold. Based on our analysis, the pulse strategy reduces the number of curing operations by 36.8% compared to traditional asynchronous curing strategies in the Markovian SIS model.

All the above-mentioned theoretical analyses are verified by directly simulating SIS processes.

Samenvatting

Verspreidingsfenomenen zoals de verspreiding van ziekten, informatie en computervirussen zijn alomtegenwoordig in natuurlijke systemen en systemen die door mensen zijn gemaakt, maar ons begrip hierover is nog altijd onvoldoende. Dit proefschrift richt zich op de analyse van het wiskundige basismodel van verspreidingsprocessen over onderliggende netwerkstructuren en heeft tot doel om de basistheorie over verspreidingsprocessen te completeren. We verkennen het *Susceptible-Infected-Susceptible* (SIS) virusverspreidingsmodel vanuit meerdere relevante perspectieven om een bijdrage te leveren aan het begrip van het model.

Onze eerste en belangrijkste bijdrage is gerelateerd aan het probleem van temporale correlaties. In de meeste studies is de invloed van tijd in het SIS verspreidingsproces weggelaten omdat tijd de specifieke waarde van eerstemoment metastabiele eigenschappen zoals de besmettingskans van een node niet beïnvloedt. Alleen de ratio van de herstel- en infectiegraad doet er toe. In dit proefschrift laten we zien dat de temporale correlatie analyseerbaar is met gemiddeldveld benaderingen ondanks dat gemiddeldveld methoden alleen bedoeld zijn om de eerste-moment eigenschappen te analyseren. We leiden de autocorrelatie van de nodale infectiestatus zowel in de stabiele als in de transiënte toestand af onder de gemiddeldveld benadering. Door de autocorrelatie in het SIS-proces te analyseren met een gemiddeldveld benadering indiceren we de invloed van het onderliggende netwerk en de waarde van de herstel- en infectiegraad op tijdgerelateerde eigenschappen van het verspreidingsproces. We laten ook zien dat de infectie- en herstelgraad te berekenen zijn door de infectiestatus van elke node te meten.

Ten tweede rekken we de Markoviaanse aanname in het SIS proces op door uitbreiding van het Poisson infectieproces naar een Weibull vernieuwingsproces. Het Poisson infectieproces is slechts een speciaal geval van het Weibulliaans vernieuwingsproces. We focussen specifiek op de extreme (limiterende) casus van het Weibulliaanse SIS proces waarin de verdeling van de infectietijd een Dirac deltafunctie is. In dit Weibulliaans kader kunnen we het niet-Markoviaanse infectiegedrag parametriseren en enkele nieuwe kenmerken laten zien die hieruit voortkomen. De analyse van de extreme casus leidt tot de grootst mogelijke epidemische drempelwaarde voor niet-Poissoniaanse infectieprocessen. Vervolgens bediscussiëren we de epidemische drempelwaarde voor verschillende infectieprocessen met Weibull-, lognormaal- en gamma-verdeelde infectietijden, welke in lijn zijn met realistische verspreid-

ingsfenomenen volgens een eerdere, op vernieuwingstheorie gebaseerde, niet-Markoviaanse gemiddeldveld methode. We laten zien dat onze resultaten in overeenstemming zijn met eerder opgestelde theorie en dat deze verschillende infectieprocessen zich op dezelfde manier gedragen.

Ten derde bestuderen we localisatiefenomenen in netwerken vanuit het gezichtspunt van SIS-verspreidingsprocessen. In netwerken waarvan de eigenvector van de adjacency matrix met de grootste eigenwaarde is gelocaliseerd kan het verspreidingsproces net boven de epidemische drempelwaarde worden gelocaliseerd. In het geval van gelocaliseerde verspreiding convergeert de prevalentie (het verwachte aantal infecties) naar nul bij toenemende netwerkomvang terwijl het aantal geïnfecteerde nodes niet nul is. Dus gelocaliseerde verspreiding vormt een interessante fase die zich onderscheidt van de fase zonder infectie en de endemische fase met een incidentie ongelijk aan nul. We evalueren de bovengenoemde extreme casus van het Weibull SIS proces waar de tijd-afhankelijke prevalentie op lange termijn periodiek is. Dichtbij de epidemische drempelwaarde divergeert in sommige netwerken de ratio van de stabiele toestand maximum en minimum incidentie die gelijk is aan de grootste eigenwaarde van de adjacency matrix, maar het verspreidingsproces is nog steeds gelocaliseerd. In andere woorden, de divergerende prevalentieratio, die wordt bepaald door de grootste eigenwaarde van het netwerk, kan een prevalentie met waarde nul niet versterken naar een prevalentie ongelijk aan nul in de thermodynamische limiet. Het resultaat indiceert dat de localisatie van verspreidsingsprocessen alleen wordt bepaald door de structuur van het netwerk maar niet door het specifieke infectieproces.

Tot slot bestuderen we de genezingsstrategie voor de beheersing van het verspreidingsproces en in het bijzonder de *pulse curing* strategie. In tegenstelling tot een klassieke asynchrone genezingsstrategie (bijvoorbeeld een Poissoniaanse) is de *pulse curing* strategie een geoptimaliseerde methode die verspreiding onderdrukt en deze strategie wordt breed toegepast in de ziektebeheersing. We hebben een model bestudeerd dat bestaat uit een *susceptible-infected* proces en een periodiek *pulse curing* proces met een genezingskans kleiner dan een. We hebben de gemiddeldveld epidemische drempelwaarde afgeleid. Gebaseerd op onze analyse blijkt de *pulse curing* strategie het aantal genezingsoperaties met 36.8% te reduceren ten opzichte van traditionele asynchrone genezingsstrategiën in het Markov SIS model.

De bovengenoemde theoretische analyses zijn geverifieerd door middel van directe simulaties van SIS-processen.

1 Introduction

"I simply wish that, in a matter which so closely concerns the wellbeing of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide."

—Daniel Bernoulli, 1760.

Fitzgerald: "The rich are different from us."
Hemingway: "Yes, they have more money."[1]

1.1 Modelling spreading phenomena

SPREADING phenomena are ubiquitous in nature, society, and technical systems, such as the propagation of diseases, computer virus, faults, and news. Understanding spreading phenomena relies on mathematical modelling since conducting a spreading experiment is difficult especially in large scales and the observation of real spreading processes is hardly detailed enough. As early as 1760, Bernoulli [2, 3] analyzed the mortality of spreading disease smallpox. His analysis may be the first known mathematical epidemic model and as a pioneer, Bernoulli seems to believe the power of mathematics in modelling reality as indicated by the above-mentioned quotation. Modern mathematical epidemic models are based on the theory proposed by Kermack and McKendrick [4, 5], which divides the population into compartments such as susceptible, infected and recovered and the spreading becomes dynamical transformations among the compartments. Of course, those compartments can be interpreted differently in contexts other than infectious diseases. The math-

[1] A dialogue in imagination quoted by P. W. Anderson in his paper 'More is Different' [1].

ematical models are powerful in understanding spreading phenomena, even only incorporating a homogeneously mixed population or a lattice [6, 7] where each component of the system plays the same role. However, a spreading may not run on a uniform or well-mixed medium. Behind the spreading dynamics, underlying structures are bearing and guiding the spreading trajectories. Historically, people noticed this when evaluating the statistical data of diseases. During the 1854 Broad Street cholera outbreak, by marking the death cases on the map of London, John Snow, who believed the cholera is caused by contaminated water, found a spatial correlation between the prevalence of cholera and a water pump and also a correlation between the prevalence and a water supply company which provided water taken from polluted sections of the River Thames [8]. Those correlations confirm the conjecture that the cholera is not spreading through the air but water as the underlying structure influences the spreading mode [9].

During the past twenty year's development of network science, people realize that real systems, such as social networks and the internet, exhibit many non-trivial features that cannot be captured by simple homogeneous models. Real systems represented by networks are usually extremely heterogeneous; they are ordered and random at the same time; they are very large but the diameter can be extremely small. Thus, the interplay between the underlying network structure and the dynamics on it becomes an interesting topic [10, 11]. Incorporating those complex networks with the compartmental epidemic models reveals new understandings about the real spreading phenomena. The seminal work [12] of Pastor-Satorras and Vespignani indicates that the epidemic threshold of a scale-free network with a finite average degree [13] may tend to zero with the increase of the network size, which is a bad news for the disease control: No matter how much effort is investigated, there may be still an epidemic of disease since the global social network is so large. Later, it is understood that just above the epidemic threshold the epidemic process is actually in an inactive Griffiths phase [14, 15, 16] or a localized spreading state [17] where the fraction of infected nodes tend to zero even though the number of infected nodes is not zero. Those results may liberate us from the fear of predestined global epidemics.

In a word, the mathematical study of the spreading processes indeed refreshes our understanding and provides a lot of insights just as Bernoulli believed.

1.2 Complexity of spreading

No doubt, the network structure influences the behavior of the spreading, but on the other hand, the spreading process is also a probe to explore the network properties. The study of the networked spreading does not only help us to understand the real spreading phenomena but also make it possible revealing the mystery that how simple local rules lead to global emergent behavior in complex systems.

What is the complexity of an object? One may have different definitions for different objects. When an object is divided into several (or a large amount of) simpler parts, we should expect that the complexity of the object is reduced. For example, nodes in a heterogeneous network may play very different roles, but the network can be dismantled into similar nodes with a similar number of links connected to them. Anderson indicates that *more is different* [1]: The whole is different from the sum of its parts. If the whole is complex, then the whole must have something new compared to its simple parts. In other words, the complexity is about how new the combination of simple components is. We then also notice that simplicity and complexity are relative concepts, because a simple component may be a complex object from the viewpoint of its lower-level components.

Based on the arguments above, we may exam whether and when an object is complex. Apparently, human society is complex because society has new cultural features and we cannot study human culture by only studying each biological person [18]. The overall distribution of human height is simple: It is possible to know the overall distribution of the human body's weight or height[2] by studying an individual since biologically people are independent and similar. The correctness of the overall estimation is guaranteed by the law of large numbers. It might be true that to form complexity, the law of large numbers should not be present and equivalently, interactions between a large number of components are mandatory.

A networked spreading process may be complex since the underlying network is composed of a large number of nodes and there are interactions between neighbors, i.e. an infected node can infect its healthy neighbors in the language of epidemiology. The infection rate between the neighbors characterizes the strength of the interaction. Now, something new comes out: If we increase the infection rate from zero, then at a certain value, the spreading pro-

[2]Even though people from some countries, e.g. the Netherlands, are taller than the others generally.

cess suddenly becomes persistence on the network. The critical value is the epidemic threshold determined by the network structure. If the infection rate is below the epidemic threshold, then the spreading dies out in the long run. If the infection rate is above the value, the fraction of infected nodes fluctuates slightly around its average. One will wonder what happens if the infection rate is exactly at the epidemic threshold. At the threshold, the fluctuation becomes very large and the number of infected nodes will follow a power-law distribution [19], which means that the infection can happen at any scale of the network compared to the situation where the infection rate is far from the epidemic threshold. This is a typical critical phenomenon and criticality means that the spreading is just located at the edge between two phases: The endemic phase and the all-healthy phase. The historical data indicates that the distribution of the blackout size follows a power-law tail [20] and if we consider the blackout as a spreading process, the power system seems indeed working at the critical point. The blackout can happen at a very large scale and influence millions of people, though not frequently. The disease spreading can also happen at a wide variety of scales up to worldwide. During 1918 and 1919, the Spanish flu spread across Europe, America, and Asia, reaching to Micronesia and the North Pole and caused 500 million infections and 50-100 million deaths [21]. Again, we do not see a global disease spreading so often but it sometimes does happen.

If the real spreading process works at the critical point, how could it be possible that the infection rate is fine-tuned to the epidemic threshold and what the advantage of working critically? For the first question, theories of self-organized criticality try to explain those phenomena, such as the sandpile model [22]. In the spreading model, due to the existence of the Griffiths phase or localized infection as mentioned in section 1.1, the critical point of the spreading process is replaced by a critical region in complex networks [19] and the behavior of spreading processes in this region is similar to that just at the critical point. The critical region does not exist in simple regular graphs but exists in complex networks that model real systems. Thus, spreading processes in nature do not need to be fine-tuned to exact critical points. For the second question, different systems may lead to different answers. If we consider the brain system and assume the activation of neurons is a spreading process, then working at the critical region saves the energy to keep the minimum number of neurons active while takes the advantage of the whole brain since the spreading can happen at any scales of the brain[3].

[3]Recently, the theory of the SIS Griffiths phase is applied to evaluate whether or not life can exist in two spatial dimensions: 2D life may need planar graphs which experience SIS Griffiths

The above two sections are too short to cover the issues about the history of mathematical modelling and the complexity of spreading. However, through this short review, we may sense that the study of spreading processes is not only useful but also fascinating. The research presented following aims to contribute new understandings about spreading phenomena.

1.3 Motivation and research questions

We study the spreading phenomena by investigating a specific epidemic model, the susceptible-infected-susceptible (SIS) process on networks. Although extensive studies have been conducted, the understanding of the SIS model is still insufficient. In this dissertation, we aim to understand the following aspects.

The first aspect is about the temporal correlation of the SIS process. Most of the studies focus on the metastable state of the process where the infection probability of each node is almost constant. Given the underlying network, the infection probabilities and the phase transition of the process are only determined by the effective infection rate, i.e. the ratio between the infection rate and the curing rate. However, the value of those rates coupled with the network structure determines the speed of the dynamical evolution. Calculating the autocorrelation of the infection state of each node is a way to understand the time-dependent behavior. Since the mean-field methods only omit the correlations between different nodes, it is possible to use them to calculate the correlation of the infection state of the same node at different time points. Furthermore, since the rates contain information about the changing speed of node state, it is also possible to calculate the value of the infection and curing rates given the observation of time-dependent node state.

The second aspect is about the limitation of the Markovian assumption in traditional studies and how the epidemic threshold changes if the SIS process is not Markovian anymore. The SIS process is Markovian in the sense that the infection and curing processes are assumed to be Poisson processes and the exact SIS model can be represented by a Markov process. The mean-field methods are approximations of the Markov process. If the infection or curing is not Markovian anymore, the mean-field approximations become invalid. It has been shown that a non-Markovian assumption can alter the epidemic threshold largely. What we are interested in is how the epidemic threshold changes for general non-Markovian spreading processes.

phase as their neural networks. Scargill [23] shows that such kind of planar graphs exist.

The third aspect is related to the localization of the spreading process. The localization phenomena appearing in different dynamical processes on networks are due to the localization of the principal eigenvector of the adjacency matrix of the network which governs the probability distribution of the system state. In an SIS process just above the threshold, the infection probability of each node is proportional to the value of the corresponding element in the principal eigenvector. Thus, the spreading is also localized. However, if a special infection process allows the infection probability non-constant and the ratio between the maximum and minimum probability diverges, could the localized spreading becomes a delocalized one? Does the localization of the spreading process depend on the infection process?

The last aspect is about the curing strategy. The pulse curing is broadly applied in public health. This strategy seems to be an optimal one because synchronized curing of all nodes in a network can immediately shut down the spreading no matter how fast the virus spreads. How to quantify the efficiency of the pulse strategy? Does the efficiency depend on the specific network structure?

1.4 Dissertation organization

The dissertation is organized as follows. In chapter 2, we define the SIS process on networks and introduce the mean-field approximation to analyze it. The simulation of the SIS process is discussed. Apart from the basis, we discuss the inaccuracy of the mean-field calculation due to the initial conditions and show the non-unimodal properties of the prevalence due to the network structure. In chapter 3, we study the autocorrelation of the infection state both in the transient and metastable state. In chapter 4, we extend the Poissonian infection process to a Weibull renewal process. The limiting case of the Weibull distribution is evaluated by time-dependent governing equations and the possible range of the epidemic threshold of non-Markovian SIS processes is conjectured. Moreover, we indicate that for other infection processes with lognormal and Gamma time distributions, the SIS process behaves similarly. In chapter 5, the localization of SIS processes is explored by looking into the limiting case studied in chapter 4. By evaluating the SIS process near thresholds, we emphasize the role of the network structure in the localized spreading. In chapter 6, a short analysis of the pulse curing strategy is provided. We quantify the effect of the pulse curing by comparing the epidemic threshold between the pulse curing and the uniform curing.

2

Spreading Process on Networks

In this chapter, we briefly review the complex networks and introduce the SIS model which we focus on throughout the dissertation. The mean-field analytic tools and simulation methods are explained. Finally, we indicate the inaccuracy of mean-field approximations due to initial conditions and show a new phenomenon of the network SIS model comparing to a well-mixed model, i.e. non-unimodality of the prevalence.[1]

2.1 Complex networks

COMPLEX networks model the interacting relationships among components in complex systems [27], such as power grids [28, 29, 30, 31], the Internet [32], social [33, 34, 35, 36] and biological systems [37, 38, 39, 40]. The complex networks are large graphs but are not purely regular or random, usually characterized by short average path length [41, 42, 43], high clustering [44], community structures [45], assortative mixing [46] and heavy-tailed degree distributions [13].

Network models also provide insights in applications, such as the design of artificial neural networks [47, 48] and error-correcting codes [49, 50]. Nowadays, data obtained from real-world features complex correlation structures. Network science is more than just conceptually grasping the characteristics of reality but advancing the processing of complex data, such as community detection [51, 52], network estimation [53, 54, 55, 56], machine learning on graphs [57], and graph signal processing [58].

[1] This chapter is partially based on [24, 25, 26].

2.1.1 Brief review of graphs and networks

The city Königsberg in Prussia was divided into two mainland portions and two large islands which are connected by seven bridges. Euler realized [59, 60] that there is no solution to the problem: Walk through all bridges by crossing each bridge once and only once. To analyze the problem rigorously, Euler abstracted the islands and mainland portions as *nodes*[2] and the nodes are connected by *links* which represent the bridges. The nodes and links form a *graph*. The number of links connected to a node is the *degree* of that node. If any two nodes communicate through links, then the graph is connected. Euler found that if there is a way to walk through all the links once and only once, then the graph should be connected and the total number of nodes with an odd degree is either 0 or 2. The reason is that the walker must start from one odd-degree node and end up to another odd-degree node if there exist odd-degree nodes. Euler's study of the Königsberg seven-bridge problem initialized the modern graph theory which focuses on properties of graphs and has a broad application in different fields.

A graph models the relationships of a group of objects and the relationship between two objects can be stochastic. Erdös and Rényi introduced the concept of random graphs by combining graph theory and probability theory. They studied a model which is afterward called Erdös-Rényi (ER) graph [61]. In an ER graph, every node pair is connected by a link with a probability p. An interesting property of the ER graph is that if p is larger than a certain threshold p_c, then the graph is connected with probability 1 when the total number of nodes goes to infinity. For a fixed number of nodes, the random graphs usually follow the Boltzmann distribution [62]. The ER graph well reflects the small-world property of many systems because the average path length from one node to another is small in ER graphs [63]. The small-world property has been noticed, especially for social networks, for a long time. The Hungarian writer Frigyes Karinthy[3] in his 1929 article *Chain* [64] described the phenomenon that any two persons on the Earth can be connected by no more than five acquaintances, which is now known as the six-degree separation. Stanley Milgram [65] conducted a social experiment of counting the number of forwards of mails within 300 volunteers and found that most mails reaching the target go through around 6 intermediaries.

However, real networks [66, 67, 68, 69] usually have a high clustering

[2] In modern terms.

[3] One can find his photo in the central cafe Budapest where he visited often to write his novels.

coefficient[4]. The neighbors of a node have a higher chance to be connected than those nodes without common neighbors, which is not captured by the ER graph: The ER graph has a low clustering coefficient. Watts and Strogatz [44] found that random graphs between the regular graph and ER graph have both small average path length and large clustering coefficient. A regular graph that is highly clustered has a large average path length and by rewiring each link in the graph randomly the average path length decreases exponentially fast. In an experiment, Watts and Strogatz [44] show that the average path length approximately equals to the corresponding ER graph when only one percent of the links are rewired, but the clustering coefficient remains as high as the regular graph during the rewiring.

The Watts-Strogatz (WS) model indeed captures the collective dynamics of small-world networks but the degree distribution of the WS network is inconsistent with most real networks. Each node in the WS model has the same expectation of degree and all nodes are more or less equal. However, in real networks, a few nodes have a very high degree playing a role as a hub and those nodes are more important than others [70]. In a log-log plot, the degree distribution of many real networks [71, 32, 67, 68, 69, 72] is approximately a straight curve over several orders of magnitude and thus follows a power-law [73] distribution $\Pr[D = d] \sim d^{-\gamma}$. Barabási and Albert [13] indicated that the power-law networks can be generated by a preferential attachment mechanism.

In the Barabási-Albert (BA) model, the network is generated in a growing way. A newly arrived node is connected to the existing nodes but with a probability proportional to the degree of the existing nodes[5]. In the thermodynamic limit[6], the generated networks suppose to have a power-law exponent $\gamma = 3$. Variations of power-law networks were proposed [75, 76, 77, 78] besides the BA model. Power-law networks also have small average path length [79]. Power-law networks are also called scale-free networks since the scaling relationships among the probabilities of degrees are invariant to the specific values of those degrees and networks with different size looks similar.

[4]If a node has a degree d_i, then the number of node pairs among the node's neighbors is $d_i(d_i - 1)/2$. The clustering coefficient of the node is the ratio between the number of actually existing links among neighbors and $d_i(d_i-1)/2$. For the whole graph, the clustering coefficient can be defined as the average of the node clustering coefficient over all nodes.

[5]If the node is connected to the existing nodes with a constant probability, then the network follows an exponential degree distribution [74] in which the probability of existing a hub is smaller than BA model.

[6]The network size goes to infinity.

Although a large number of studies confirm that power-law networks are ubiquitous, it is still controversial [80, 81] that whether real networks are following a power-law degree distribution or not. Seems that most real networks cannot pass the statistical test of power-law distribution even for the network realizations generated by BA model [80, 82]. However, the controversy could be just a result of a different understanding of the power-law concept between the complexity science and data science [83].

The spreading processes are evaluated on different networks in later chapters. The interplay between the spreading dynamics and the underlying network structures is the main gradient of the dissertation. One can find more detailed introductions to graphs and complex networks in related books [74, 84, 85, 86, 87, 88, 89].

2.1.2 Adjacency matrix

The adjacency matrix will appear frequently through this dissertation. An N-node network can be represented by the adjacency matrix $A = [a_{ij}]_{N\times N}$, where a_{ij} is the weight of the link from node i to node j. In most cases, we deal with simple networks which are connected, undirected, unweighted networks without self-loops. For simple networks, each element in the adjacency matrix A follows $a_{ij} = a_{ji} = 1$ if node i and j are connected. If node i and node j are disconnected, then $a_{ij} = a_{ji} = 0$. Since no self-loop exists, $a_{ii} = 0$.

For simple networks, the adjacency matrix A is a real symmetric matrix and A has N real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$. From Perron-Frobenius Theorem [90], the largest eigenvalue λ_1, which is also called the spectral radius, is real and non-negative. The largest eigenvalue λ_1 plays an important role in characterizing dynamical processes on networks. For example, the SIS epidemic threshold is a function [91] of λ_1. The property of λ_1 is extensively studied, such as the asymptotic value [92] and bounds [93, 94]. The spectral radius λ_1 can also be tuned by the modification of the network structure [95].

The principal eigenvector $x = [x_1, x_2, \ldots, x_N]^T$ corresponding to λ_1 can be set all positive and x is the only eigenvector whose components are all positive. The value x_i characteristics the importance of the corresponding node i, which is known as the eigenvector centrality [96]. Comparing to the simpler centrality measure, the degree, the eigenvector centrality considers the global connecting features of each node: An important neighbor of node i contributes

more to x_i comparing to less important nodes shown by the following equation,

$$x_i = \frac{1}{\lambda_1} \sum_{j \in \mathcal{N}_i} x_j$$

derived by $Ax = \lambda_1 x$, where $\mathcal{N}_i$ denotes the set of neighbors of node i. The PageRank [97] used by the Google search engine ranks web pages based on the eigenvector centrality.

The principal eigenvector is associated with many networked dynamical processes. Apart from the SIS process discussed in Chapter 5, the maximum-entropy random walker appears on node i with probability proportional to x_i in the steady state [98], while this probability is proportional to the degree in the simple random walk [99].

Besides the adjacency matrix, networks are sometimes denoted by the Laplacian matrix [90], through which a network can be mapped to a simplex in a high-dimensional space [100, 101].

2.2 The SIS process on networks

The Susceptible-Infected-Susceptible (SIS) process is a basic compartmental model in mathematical epidemiology [4] which can be used to model the spread of viruses, information, opinions, and computer malware. The SIS model is initially studied in a well-mixed manner where all individuals in the spreading system are equivalent, like those in a regular graph. Since most real systems form complex networks, it is necessary to evaluate the SIS model on networks and the SIS model indeed shows many non-trivial phenomena [11]. In this section, we first introduce how the networked SIS model is exactly defined and then introduce the mean-field approximate analysis of the model.

2.2.1 The exact Markovian SIS process

In the SIS model, each node in the network can be either infected or susceptible (healthy). The infection state of node j for $j = 1, \ldots, N$ at time t is denoted by a Bernoulli random variable $X_j(t)$: infected $X_j(t) = 1$ or susceptible (healthy) $X_j(t) = 0$. The SIS model has simple local rules that each healthy node can be infected by an infected neighbor with rate β and each infected node is cured spontaneously with rate δ. The infection and curing processes are independent and both are Poisson processes. Thus, the SIS process on an

N-node network is Markovian and can be described by a 2^N-state Markov process with one all-healthy absorbing state [102, 103]. The state transition of each node in the 2^N-state Markov process can be described as

$$\begin{aligned} X_j(t) : 0 \to 1 \quad &\text{with rate:} \quad \beta \sum_{i=1}^{N} a_{ji} X_i(t) \\ X_j(t) : 1 \to 0 \quad &\text{with rate:} \quad \delta \end{aligned} \tag{2.1}$$

Since the all-healthy state is absorbing and the network is finite, the SIS process will enter the absorbing state when $t \to \infty$. However, the SIS process can also stay in the metastable state for a long time where the infection probability of every node is almost constant if the effective infection rate $\tau \triangleq \beta/\delta$ is large enough. Generally, there is an epidemic threshold τ_c determined by the network [12, 102]. If the effective infection rate $\tau < \tau_c$, then the virus dies out quickly and every node becomes healthy. Above the threshold, the infection can persist in the network for a very long time [104]. The threshold τ_c divides the system into two different phases in the metastable state: the endemic phase and all-healthy absorbing phase. For a finite-size network, the SIS process has no sharp phase transition since the SIS process is no more than a linear Markov process.

In the SIS process, the infection probability of node j for $j = 1, \ldots, N$ follows the governing Eq. [88],

$$\frac{dE[X_j(t)]}{dt} = -\delta E[X_j(t)] + \beta \sum_{i=1}^{N} a_{ji} E[X_i(t)] - \beta \sum_{i=1}^{N} a_{ji} E[X_j(t) X_i(t)] \tag{2.2}$$

Equation (2.2) describes the exact Markovian SIS process, but higher-order moments of the infection states $E[X_j(t)X_i(t)]$ are involved in Eq. (2.2). In total, $2^N - 1$ equations are needed to solve the equation [88, p. 452] and complexity increases exponentially with network size N. Furthermore, the analysis of the SIS process is not tractable without approximation, not even for the complete graph [105].

In the SIS process, the prevalence is defined as the average fraction of infected nodes $y(t) \triangleq \frac{1}{N} \sum_{i=1}^{N} E[X_i(t)]$, and the prevalence is considered as the order parameter of the system. From a global point of view, the governing equation of the prevalence [106] can be obtained by adding Eq. (2.2) over all nodes,

$$\frac{dy(t)}{dt} = -\delta y(t) + \frac{\beta}{N} E\big[w(t)^T Q w(t)\big] \tag{2.3}$$

where $w(t) = [X_1(t), \ldots, X_N(t)]$ is the network state vector. From [90, Art. 77], the quadratic Laplacian $w(t)^T Q w(t)$, where Q is the Laplacian matrix, in (2.3) equals the number of the links which have only one end node infected at time t.

2.2.2 The mean-field analysis

The N-Intertwined Mean-Field Approximation

As Mentioned in the previous subsection, the exact SIS process is complex and usually approximations are introduced to simplify the analysis. In Eq. (2.2), higher-order terms $E[X_j(t)X_i(t)]$ are introduced. To close the differential equations, we can introduce the approximation $E[X_i(t)X_j(t)] = E[X_i(t)]E[X_j(t)]$ which is called the N-Intertwined Mean-Field Approximation (NIMFA) [102, 107]. The name *NIMFA* comes from the fact that the approximation $E[X_j(t)X_i(t)] = E[X_j(t)X_i(t)]$ is equivalent to approximating the infection rate $\beta \sum_{i=1}^{N} a_{ji} X_i(t)$ of a healthy node in the process (2.1) by its average $\beta \sum_{i=1}^{N} a_{ji} E[X_i(t)]$, and the exact SIS process is approximated by N intertwined Markov processes. NIMFA is sometimes called Quenched mean-field approximation (QMF) [108], or Individual-based mean-field approach (IBMF) as in [11].

For Bernoulli random variables, uncorrelation $E[X_i(t)X_j(t)] = E[X_i(t)]E[X_j(t)]$ and independence $\Pr[X_i(t), X_j(t)] = \Pr[X_i(t)] \Pr[X_j(t)]$ are equivalent [109, footnote 5]. In the exact SIS process, the infection states of neighbors are always positively correlated [110, 111, 112, 113], i.e. $E[X_i(t)X_j(t)] > E[X_i(t)]E[X_j(t)]$.

Under NIMFA, the governing equation of is

$$\frac{dv_j(t)}{dt} = -\delta v_j(t) + \beta \left(1 - v_j(t)\right) \sum_{j=1}^{N} a_{ji} v_i(t) \tag{2.4}$$

where $v_j(t)$ is the NIMFA infection probability of node j at time t and $v_j(t)$ approximates the exact infection probability $E[X_j(t)]$. The NIMFA epidemic threshold is $\tau_c^{(1)} \triangleq \frac{1}{\lambda_1}$, where λ_1 is the largest eigenvalue of the adjacency matrix A and the superscript (1) refers to the fact that NIMFA is a first-order approximation. If the effective infection rate $\tau > \tau_c^{(1)}$, then the infection can persist on the network and the steady-state infection probability $v_{j\infty} \triangleq \lim_{t\to\infty} v_j(t) > 0$ is constant [114, 102] for non-zero initial state. The

steady state of NIMFA corresponds to the metastable state of the exact SIS process. If $\tau < \tau_c^{(1)}$, then the infection decays exponentially fast toward the all-healthy state $v_{j\infty} = 0$. Diffrent from that in the exact process, the NIMFA equations are non-linear and a sharp phase transition appears for finite-size networks. The NIMFA threshold is a lower bound of the exact threshold [115] and the infection probability $v_j(t)$ is an upper bound of the exact probability $E[X_j(t)]$ due to $E[X_i(t)X_j(t)] > E[X_i(t)]E[X_j(t)]$. As a result, the NIMFA prevalence $y^{(1)}(t) \triangleq \frac{1}{N}\sum_{i=1}^{N} v_i(t)$ is also an upper bound of the exact prevalence $y(t)$.

To avoid ambiguity, we denote the NIMFA infection state of node j at time t by another Bernoulli random variable $V_j(t)$: infected $V_j(t) = 1$ and susceptible $V_j(t) = 0$. Thus, we can actually approximate the statistical properties of the infection state $X_j(t)$ by those of $V_j(t)$ in NIMFA. In the steady state $t \to \infty$ and $dv_j(t)/dt = 0$ for $j = 1, \ldots, N$, we denote the infection state of node j by $V_{j\infty}(t) \triangleq \lim_{t\to\infty} V_j(t)$. Under NIMFA, the transition of the infection state $V_j(t)$ of node j following Eq. (2.4) can be denoted by a two-state Markov process [107], and the transition rate of $V_j(t) : 0 \to 1$ becomes a determined function of time. As mentioned, the whole system is composed of N intertwined 2-state Markov processes instead of being a 2^N-state Markov process. Corresponding to (2.1), the transition of the NIMFA infection state $V_j(t)$ is

$$\begin{aligned} &V_j(t) : 0 \to 1 \quad \text{with rate:} \quad \tilde{\beta}_j(t) \triangleq \beta \sum_{j=1}^{N} a_{ji} v_i(t) \\ &V_j(t) : 1 \to 0 \quad \text{with rate:} \quad \delta \end{aligned} \tag{2.5}$$

The infinitesimal generator of the Markov process (2.5) is

$$Q_j(t) \triangleq \begin{bmatrix} -\tilde{\beta}_j(t) & \tilde{\beta}_j(t) \\ \delta & -\delta \end{bmatrix} \tag{2.6}$$

In this dissertation, theoretical analyses of the Markovian SIS process is mainly based on NIMFA.

Other approximate methods for SIS processes

Apart from NIMFA, other mean-field methods are adopting different approximations and forms of moment closure [116]. The Heterogeneous Mean-Field (HMF) method analyses the SIS process on a network with a given degree distribution [12]. HMF additionally assumes that the infection probabilities of

nodes with the same degree are equal and thus HMF performs worse compared to NIMFA, which takes advantage of the adjacency matrix [117]. HMF has also been extended to networks where the degree correlations between nodes are considered [118]. HMF and the moment closure method [116] are also special cases of an approximate master equation for binary dynamics [119, 120]. There are also second-order approximations by incorporating higher-order moments of the infection state [121, 122]. Percolation theory and message passing, which are usually applied in non-recurrent state-changing processes, e.g. SIR [123, 124, 125], are also available in the study the SIS process [126, 127]. Furthermore, discrete-time version approximation leads to some similar understandings on network SIS model [128, 129].

Before the concept of network science [85], spreading processes on network structures have been explored in the field of mathematical biology or epidemiology. For example, a Gonorrhea spreading model [130] designed for a non-homogeneous population is a mean-field model. In the model, the relationship between the equilibrium and the principal eigenvalue λ_1 of the contact structure is studied. The correlation of infection states which is usually omitted in approximations has also been discussed [113] before the popularity of network science.

2.3 Simulation of the SIS process

Apart from mean-field analysis, direct simulation is another indispensable path of studying the SIS process. Simulation results provide an intuitive understanding and exact values comparing to mean-field theoretical results. New phenomena can be found by playing with the simulator, which also promotes theoretical analysis. One example is the oscillating prevalence shown in Fig. 4.2 which is first found by simulations and then leads to a theoretical analysis in Chapter 4 and the study in Chapter 5. Simulations might be also the only method when studying some aspects of the epidemic process when there are no proper theoretical tools, such as the Griffiths effects [15].

Since the SIS epidemic process is just a Markov process, the exact simulation can be implemented based on the Gillespie Algorithm [131, 132, 133, 134] and a brief review of the simulation can be found in [135].

In this dissertation, the simulation is done by the SIS Simulator (SISS) [136] implemented by Java. In the SISS simulator, the infection and curing events of each node are recorded on a timeline. 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 timeline, 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 timeline 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 timeline 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 timeline, because before neighbor k can be infected, node j has already been cured and lost the ability of infection. After the infection event is inserted into the timeline, a new infection event is generated again and compared with the remaining curing time. At time $t = 0$, initially infected nodes should be selected and thus the simulation starts.

The SISS is an event-based simulation method and the inter-arrival time of infection and curing can be changed to other distributions to simulate non-Markovian SIS processes. The above-mentioned events, i.e. infection and curing, are processed in an order according to the time. To search the most recent event waiting for processing, SISS maintains all the generated events in a binary-tree structure and the complexity at each update step is $O(c \log N)$, where N is the network size and c is related to the average degree [135]. The computational complexity is higher in a dense network than in a sparse network. Another available event-based simulator implemented by Python can be found in [137, 138] whose results are identical to SISS [7]. There are variant implementations apart from the event-based method [134]. St-Onge *et al.* [135] show that the complexity of each update can be reduced to $O(\log \log N)$ by a composition and rejection based algorithm.

2.4 Inaccuracy of mean-field analysis due to die-out

As the correlation of infection state between neighbors is always non-negative [110], the mean-field infection probability $v_i(t)$ is always an upper bound of the exact infection probability $E[X_i(t)]$. NIMFA is found performing well on dense networks [139] or heterogeneous networks. Besides the net-

[7] The results are compared in personal communication with J. Miller.

work structure part, the mean-field can be inaccurate due to the non-omittable die-out.

Since the non-zero solution $v_{i\infty}$ of NIMFA is globally stable [114], any non-zero initial state will asymptotically evolve to the non-zero solution $v_{i\infty}$ if the effective infection rate τ is above the epidemic threshold $\tau_c^{(1)}$ while the exact SIS process will eventually enter the all-healthy absorbing state for any finite-size network. Thus, the exact SIS process experiences a die-out which is not reflected in the mean-field analysis. Even though the above-threshold infection can persist on the network for a very long time in a large network [104], the exact SIS process may still experience the die-out in a relatively short time period under any one of the following two situations: 1) the number of initially infected nodes is finite; 2) The effective infection rate τ is around the NIMFA threshold $\tau_c^{(1)}$, i.e. $0 < \tau/\tau_c^{(1)} \ll 1$.

The exact prevalence $y(t)$ and the prevalence $\tilde{y}(t) \triangleq E[I(t)|I(t) \neq 0]$ counting only the non-die-out realizations of the SIS process have the relation (see Appendix A.1.1)

$$y(t) = \tilde{y}(t)(1 - \Pr[I(t) = 0]) \tag{2.7}$$

Mean-field methods can only characterize the non-die-out part $\tilde{y}(t)$. If one wants to compare the mean-field results with the simulation, one needs to reduce the die-out probability due to a finite initial condition, e.g. starting from all nodes infected. In the study of near-threshold behaviors, it might be better discarding the die-out realizations as the simulation results (especially Fig. 5.5) in Chapter 5.

Here we discuss a bit more about the early die-out due to the finite initial conditions. We find by simulations that the following formula approximates the die-out probability before metastable state well if the infection rate τ is above the epidemic threshold $\tau_c^{(1)}$,

$$\Pr[I(t_m) = 0] \approx \frac{1}{x^n}, \text{ with } x > 1 \tag{2.8}$$

where t_m denotes a time point in the metastable state, $x = \tau/\tau_c^{(1)}$ and n is the number of initially infected nodes. The situation $x < 1$ is not considered since below the epidemic threshold $\tau_c^{(1)}$ the exact SIS process always die-out in a relatively short time [115]. In addition, $1/x > 1$ cannot represent a probability. When $x \leq 1$, the virus die-out probability tends to 1. If a few nodes are infected and the infection rate is above the threshold $x > 1$, then formula (2.8),

which is equivalent to $\Pr[S(t_m) = 0] \approx \mathrm{e}^{-n \log x}$, shows that the network will experience a disease outbreak, because the die-out probability decreases exponentially fast with n above the epidemic threshold ($\log x > 0$).

In the sequel, we compare (2.8) and the die-out probability $\Pr[S(t) = 0]$ obtained via simulations. The curing rate of all the calculations and simulations in this section is $\delta = 1$.

2.4.1 Early die-out in complete graphs

The Markovian SIS epidemic process on the complete graph K_N is a birth-and-death process [140, 88]. The states $\{0, 1, \cdots, N\}$ of the birth-and-death process are the number of infected nodes, where 0 is the all-healthy absorbing state. Therefore, the die-out probability $\Pr[I(t) = 0]$ can be obtained by solving the birth-and-death process,

$$\big(\boldsymbol{s}'(t)\big)^T = \boldsymbol{s}^T(t)\mathsf{Q} \tag{2.9}$$

where Q is the infinitesimal generator of the birth-and-death Markov chain, and $\boldsymbol{s}^T(t) = [s_0(t), \cdots, s_N(t)]$ is the state probability vector with each element $s_i(t) = \Pr[S(t) = i/N]$ for $0 \le i \le N$, and $s_0(t) = \Pr[S(t) = 0]$. The die-out probability $\Pr[I(t) = 0]$ of SIS epidemic process in complete graphs also equals the gambler's ruin probability [88, p. 231]. In this subsection, we show that the die-out probability is well approximated by formula (2.8) in complete graphs.

After solving (2.9) numerically for the complete graph K_{126} with effective infection rate $\tau = 0.016$, Fig. 2.1 shows the prevalence $y(t)$ and the die-out probability $\Pr[I(t) = 0]$ as an example. The metastable state is reached approximately at time $t = 10$ and hereafter the prevalence $y(t)$ keeps steady. Also, the die-out probability $\Pr[I(t) = 0]$ becomes approximate constant earlier from $t = 5$. The prevalence $y(t)$ will decrease slowly to 0 after a long time since the network is finite [141, 104], and correspondingly, the die-out probability will increase to 1. At $t = 45$ in the metastable state, the number of die-out realizations of the SIS epidemic simulation and the solution of the Markov chain (2.9) are recorded and shown in Fig. 2.2, 2.3, and 2.4. In the simulation, 10^6 realizations of the Markovian epidemic process are simulated. By counting the number of realizations which have zero infected nodes at $t = 45$, the die-out probability is obtained.

Figure 2.2 and 2.3 illustrate that, our simulation results match with the computation of the birth-and-death process (2.9). To avoid redundancy, we

omit the simulation results in Fig. 2.4. From Fig. 2.2, the die-out probability at $t = 45$ is approximately 1 corresponding to formula (2.8), when the rate $x = 1$. Also, if $x = 1$, the infection rate is below the threshold, and no matter how many nodes are infected initially, the prevalence $y(t)$ decreases exponentially fast for sufficiently large time. For a different number of initially infected nodes n, Fig. 2.2, 2.3 and 2.4 show that the virus die-out probabilities converge to the concise formula (2.8) fast with the network size N. Furthermore, the larger x is, the faster the probabilities convergence towards (2.8).

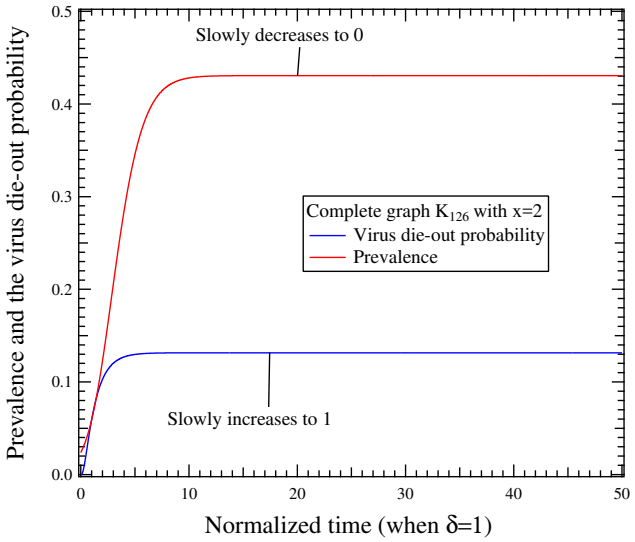

Figure 2.1: The virus die-out probability and the prevalence of epidemic process in complete graph K_{126}. Initially 3 nodes are infected. This figure shows a clearly metastable state region.

2.4.2 Early die-out in general graphs

For general graphs, it is infeasible to obtain the virus die-out probability by directly solving the differential equations of the Markov chain, because the number of equations is 2^N. However, it is still possible to obtain the virus die-out probability efficiently by simulation. We construct three Erdős-Rényi (ER) graphs $G_p(N)$ with the network size $N = 100$ and the link generation probability $p = 0.9$, 0.5, and 0.1, respectively. The epidemic process is simulated on the ER graphs by randomly choosing the initially infected nodes. For every normalized infection rate x and every number of initially infected nodes

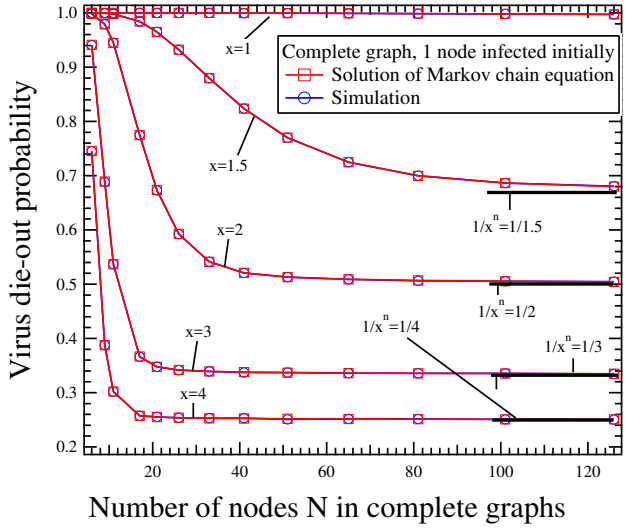

Figure 2.2: The die-out probabilities from simulation of the SIS epidemic process and calculation of the birth-and-death process are shown with $n = 1$ initially infected node. With the increase of network size N, the die-out probabilities converge to the formula: $1/x^n$.

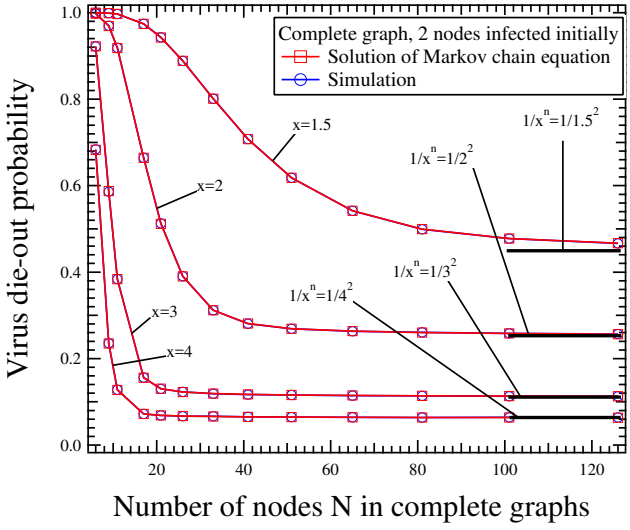

Figure 2.3: With $n = 2$ nodes infected initially, this figure verifies (2.8) as Fig. 2.2 with simulation and calculation results.

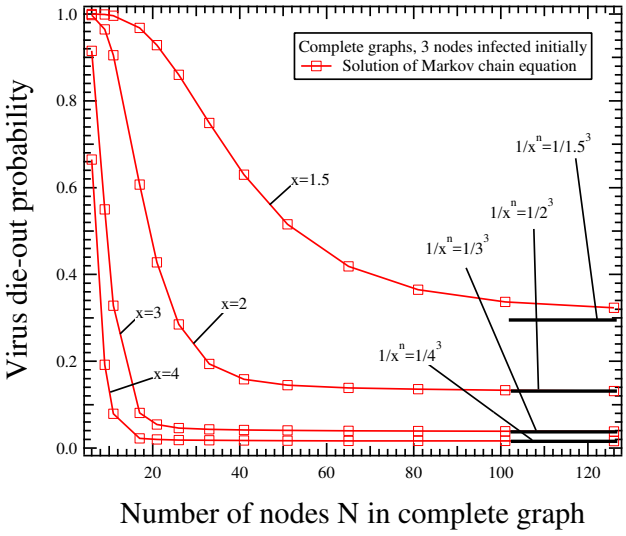

Figure 2.4: With 3 nodes infected initially, this figure shows the calculation results of (2.9) as Fig. 2.2 and 2.3.

n, 10^4 realizations are simulated. Fig. 2.5, 2.6, and 2.7 give the the comparison between the die-out probabilities and formula (2.8) for the number of initially infected nodes $n = 1, 2, 3$. Formula (2.8) is accurate in the general ER graphs, especially when the rate x is large. The accuracy of formula (2.8) decreases with decreasing link generation probability p in ER graphs $G_p(N)$.

The die-out probability of the SIS epidemic process in a power-law graph is presented in Fig. 2.8 with 10^5 realizations, and formula (2.8) shows its limitation. The power-law graph has $N = 1000$ nodes, and the degree distribution is $\Pr[k] \sim k^{-2.6}$. Figure 2.8 exhibits that the die-out probability is almost 1 when the rate x is around 2, which also indicates that the real epidemic threshold in the power-law graph is much larger than the NIMFA threshold $1/\lambda_1$. The inaccuracy of formula (2.8) is affected by the inaccuracy of the NIMFA threshold $\tau_c^{(1)} = 1/\lambda_1$ and thus $\tau_c^{(1)}$ in x might be replaced by a more accurate epidemic threshold to achieve a better approximation in the case of Fig. 2.8.

The simulations seem to indicate that formula (2.8) is always smaller than the actual die-out probability, which may be attributed to the fact that the NIMFA threshold always lower bounds the actual threshold in any network.

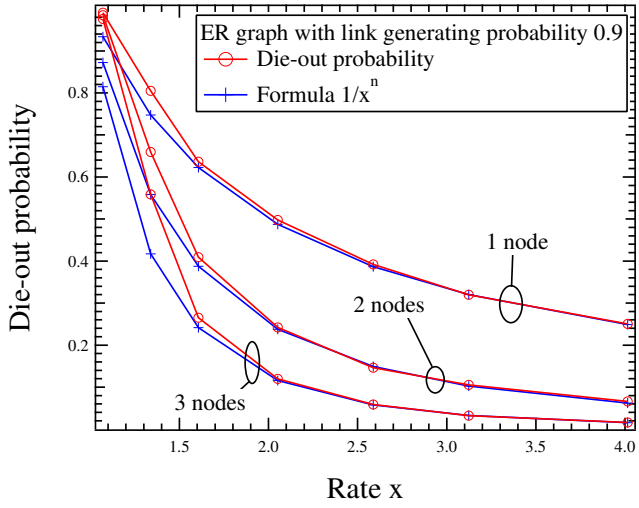

Figure 2.5: The virus die-out probability of the SIS epidemic process in an ER graph with the link generation probability 0.9. The virus spreads starts from 1, 2, or 3 nodes initially.

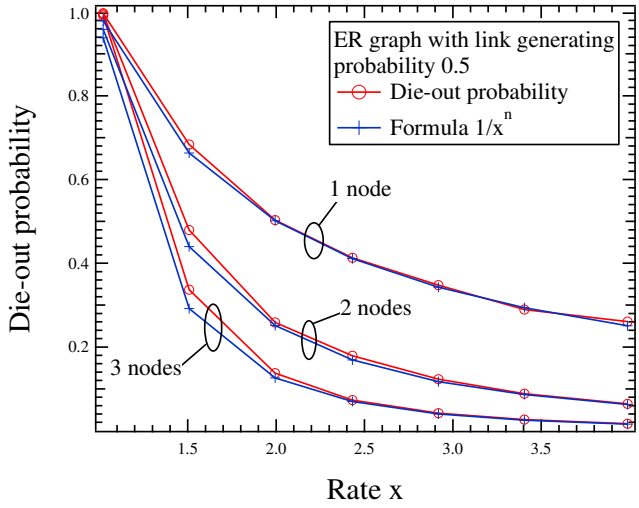

Figure 2.6: The die-out probability of the SIS epidemic process in another ER graph with the link generation probability 0.5.

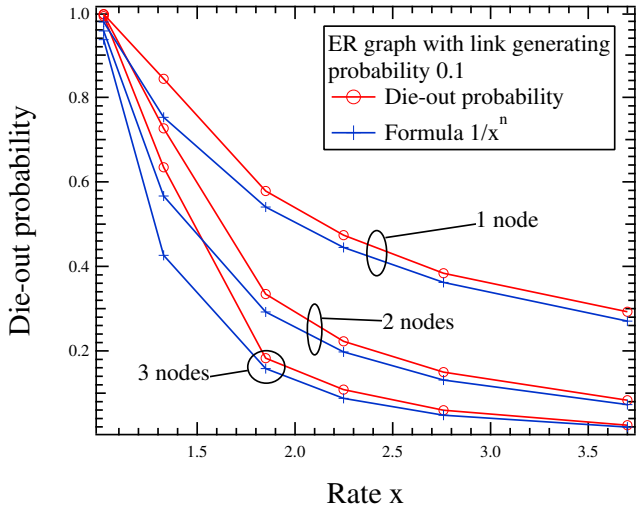

Figure 2.7: The die-out probability of the SIS epidemic process in another ER graph with the link generation probability only 0.1.

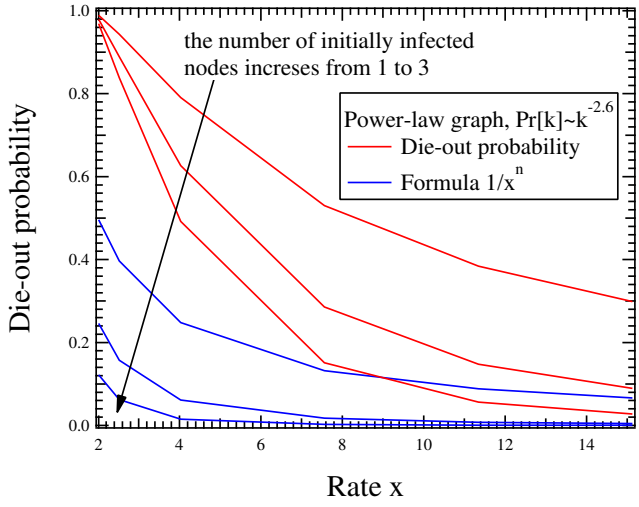

Figure 2.8: The die-out probability of the SIS epidemic process in a power-law graph.

2.4.3 Correction for die-out

When a small number of nodes is initially infected, the die-out probability is relatively large. Since we have an approximate formula (2.8) of the die-out probability and we observed in Fig. 2.1 that the die-out probability increases fast to the metastable value, we can try to correct NIMFA to fit the exact simulation results.

As we have mentioned, NIMFA is conditioned to the case where the virus in the epidemic process will not die out and the absorbing state is removed. Based on (2.8), we can approximate virus surviving probability function at the time t as

$$f(t) = 1 - \frac{1}{x^n} + \frac{1}{x^n} \mathrm{e}^{-\lambda_1 t} \tag{2.10}$$

Equation (2.10) is motivated as follows. At time $t = 0$ and $y^{(1)}(t) \neq 0$, the virus surviving probability is 1 and $f(0) = 1$. Next, simulations seem to indicate that the virus die-out probability decreases exponentially fast to $1/x^n$ in metastable state and we assume the decreasing rate is proportional to λ_1 because the infection persists easier when λ_1 is larger.

To incorporate the die-out, the NIMFA prevalence can be corrected by applying (2.7) as

$$y(t) \approx y^{(1)}(t) f(t) \tag{2.11}$$

Figure 2.9 and 2.10 present the prevalence and the approximation (2.11) of the SIS epidemic process in the complete graph K_{50} and the random generated ER graph in Sec 2.4.2. The normalized time is the time scale when the curing rate $\delta = 1$ and the prevalence is obtained by averaging 10^6 realizations. Starting from one or two infected nodes, NIMFA fails to predict the prevalence. The steady state of NIMFA is independent of the initial conditions. Fortunately, (2.11) seems a good approximation at the initial stage of the SIS epidemic process.

Due to the discussion in this section, we will let all nodes be infected when performing simulation in the study of later chapters to reduce the large early die-out probability.

2.5 Non-unimodality of the prevalence: examples

Different from the SIS model without an underlying network structure (well-mixed population), network SIS model can exhibit a non-unimodal prevalence

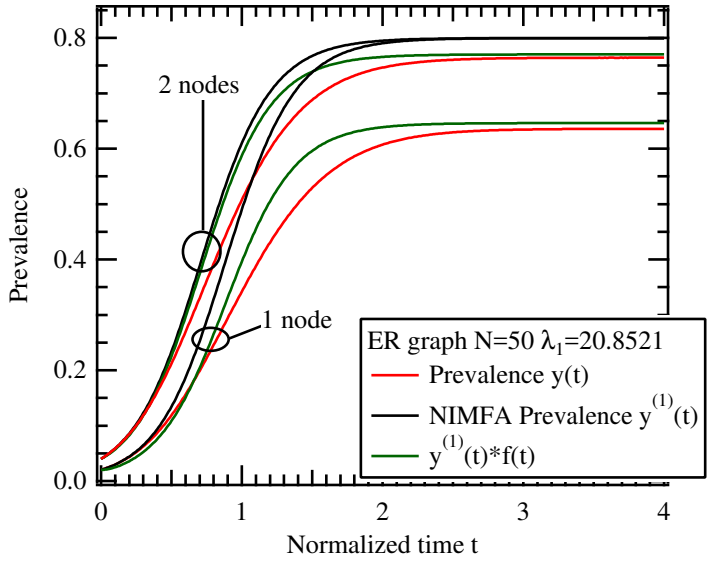

Figure 2.9: The SIS epidemic process in ER graph with network size $N = 50$. The effective infection rate is $\tau = 0.25$.

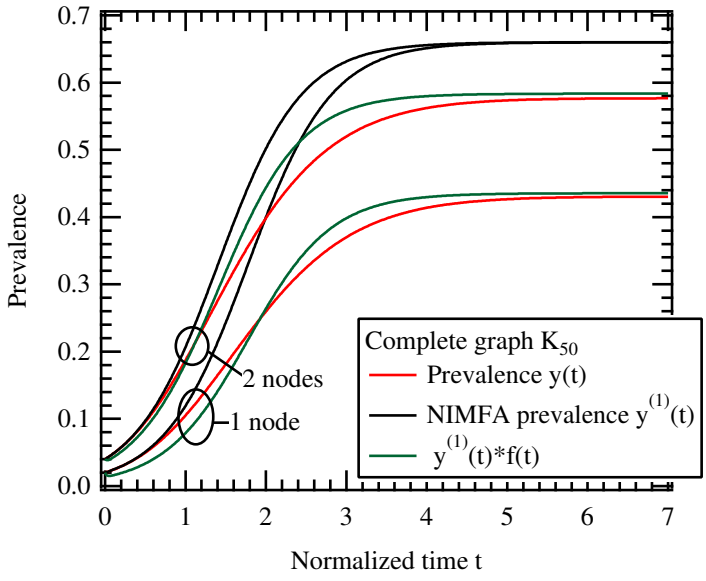

Figure 2.10: The SIS epidemic in complete graph K_{50} with effective infection rate $\tau = 0.06$.

(i.e., the prevalence has at least two local maxima) due to the variations of local network structures. Figure 2.11 shows an example of a time-dependent non-unimodal prevalence. The virus spreads starting from 600 infected nodes in a specially designed network[8] with size $N = 1740$ as shown in Fig. 2.12. Due to the absorbing all-healthy state $\lim_{t\to\infty} y(t) = 0$, the prevalence function has at least 3 local maxima for $t \geq 0$ while a unimodal prevalence has only 1 maximum [142].

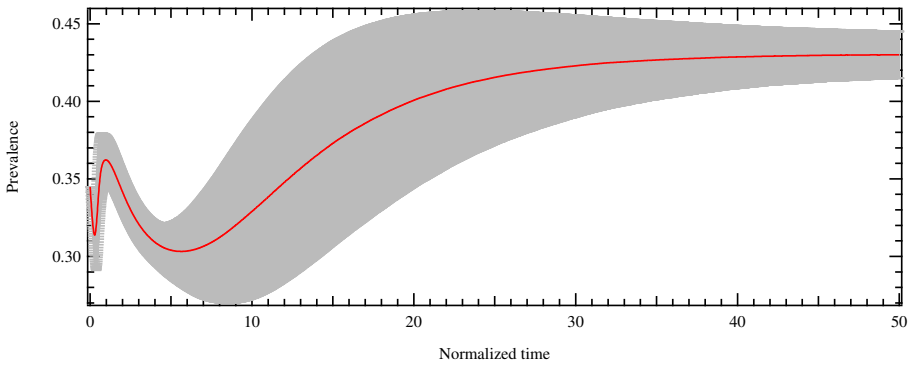

Figure 2.11: Non-unimodality of the SIS prevalence: an example with $\tau = 1$ and averaging over 10^5 realizations of simulation. The non-unimodal prevalence is shown as the red curve together with the standard deviation shown as the grey bars.

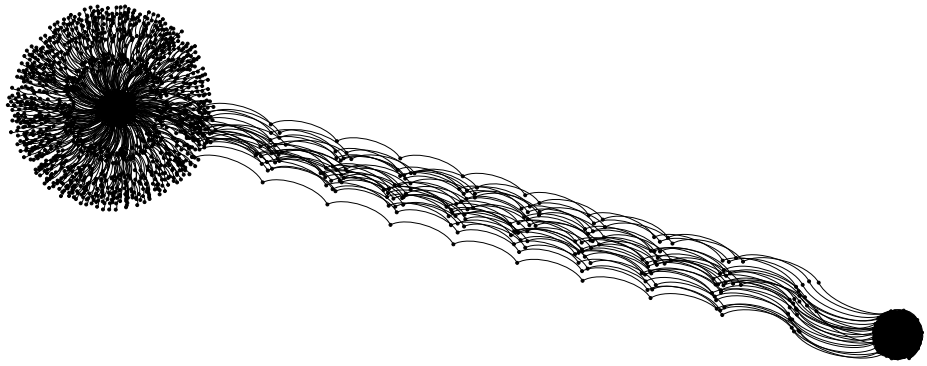

Figure 2.12: Structure of the network corresponding to Fig. 2.11

The network shown in Fig. 2.12 is consist of three parts: a 33-regular graph with size 100 in the left that each node is connected additionally with 3 short

[8]The network is provided by Joel C. Miller

4-length paths, a 44-regular graph with size 200 in the right, and several 11-length long paths connecting the two regular graphs. The virus spreads starting from the left part of the network. Initially, all two endmost nodes (totally 600 nodes) of the short paths are infected, but the infection cannot persist at the initial level with $\tau = 1$. Many initially infected nodes are cured over time and the prevalence decreases. However, the virus lately spreads to the 33-regular subgraph due to a large number of short paths, and the prevalence increases temporally as a result. After a balance of infection and curing on the 33-regular subgraph achieved, the prevalence decreases because of the dominating curing process on the short paths. Eventually, the virus spreads to the 44-regular subgraph through the long paths. The SIS process approaches the metastable state and the prevalence increases again.

The above example indicates the existence of multiple maxima of the time-dependent prevalence. Here, we consider the simplest non-unimodal prevalence function which features a decrease at the initial stage and an increase afterwards. Given the initial condition $w(0)$, the prevalence at an arbitrarily small time $\epsilon > 0$ is,

$$y(\epsilon) = y(0) + \epsilon \left.\frac{dy(t)}{dt}\right|_{t=\epsilon} + O(\epsilon) \tag{2.12}$$

If $y(\epsilon) < y(0)$, then

$$\left.\frac{dy(t)}{dt}\right|_{t=0} < 0$$

for $\epsilon \to 0$. After substituting the initial state $w(0)$ into (2.3), we have

$$\left.\frac{dy(t)}{dt}\right|_{t=0} = -y(0) + \frac{\tau}{N} w^T(0) Q w(0) < 0 \tag{2.13}$$

If we denote the initially infected subnetwork as G_I, then we arrive at

$$\tau |\partial G_I| < |G_I| \tag{2.14}$$

where $|\partial G_I| = w^T(0) Q w(0)$, and ∂G_I is the edge boundary of G_I which is the set of links[9] connecting G_I and its complementary subnetwork G_I^C. Furthermore, $y(0) = |G_I|/N$ where $|G_I|$ denotes the number of nodes in the subnetwork G_I. Conversely, if $\tau |\partial G_I| > |G_I|$, then the prevalence increases at $t = 0$. Condition (2.14) is only a necessary condition for that the prevalence is non-unimodal, because the prevalence may monotonically decrease toward 0 over time t.

[9] The cut between G_I and G_I^C

Non-unimodal prevalence is introduced by the spreading bottlenecks in the network. In the following, we show more examples of non-unimodal prevalence.

2.5.1 Two cliques having nodes in common

Two complete graphs K_N can have 1 to N nodes in common. If there are N nodes in common, then two complete graphs K_N are merged into one. If we assume that the fraction of corresponding node pairs being merged into one node is a and all nodes in one of the two cliques are infected initially, then from (2.14), the prevalence decreases initially if

$$\tau < \frac{1}{Na(1-a)} \tag{2.15}$$

As mentioned above, if $a = 1$, then the whole network is a complete graph K_N, and if $a = 0$, then the network is disconnected. In both situation $\lim_{a\to 0 \text{ or } a\to 1} 1/Na(1-a) \to \infty$, the prevalence always decreases initially no matter how large the effective infection rate τ is.

Figure 2.13 illustrates the prevalence of the Markovian SIS process on the networks with different value of a. The clique size is $N = 50$ and the effective infection rate is $\tau = 0.1$. With the fraction of common nodes a increasing, the prevalence changes from unimodal functions to non-unimodal functions. In Fig. 2.13, if $a \in (0.2764, 0.7236)$, then the prevalence is a logistic-like curve where the prevalence firstly increases and lately keeps steady in the metastable state as the red curves. If $a \in (0, 0.2764) \bigcup (0.7236, 1]$, then the prevalence decreases initially, but increases again only if $a \in (0, 0.2764)$ as the blue curves. The non-unimodal prevalence appears only if a is small. The average degree of the network is $\frac{2-a^2}{2-a}N - 1$, which reaches the maximum $\frac{7}{6}N - 1$ when $a = 0.5$. The average degree is almost constant with a different value of a, and the metastable state prevalence is on an approximately same level.

Figure 2.14 presents the NIMFA prevalence corresponding to the Markovian prevalence in Fig. 2.13. The NIMFA prevalence has a similar changing trend.

Fig.2.13 and 2.14 show three different types of prevalence. The red logistic-like curves feature a fast increase initially and then reach to the metastable state. The green ones describe that infection cannot persist at the initial level and the prevalence decreases to a reasonable level. The non-unimodal prevalence (blue curves) decreases initially and then increases

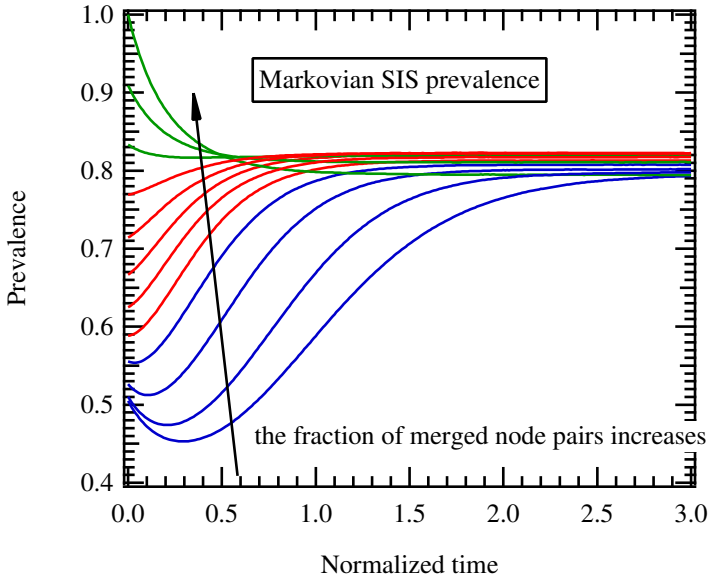

Figure 2.13: The SIS prevalence with $\tau = 0.1$ on networks which contain two cliques with size $N = 50$ with the fraction of common nodes $a = 1/50, 1/25, 1/10, 1/5$ for blue curves, $a = 0.3, 0.4, 0.5, 0.6, 0.7$ for red curves, and $a = 0.8, 0.9, 1$ for green curves. Initially, all nodes in one clique are infected initially.

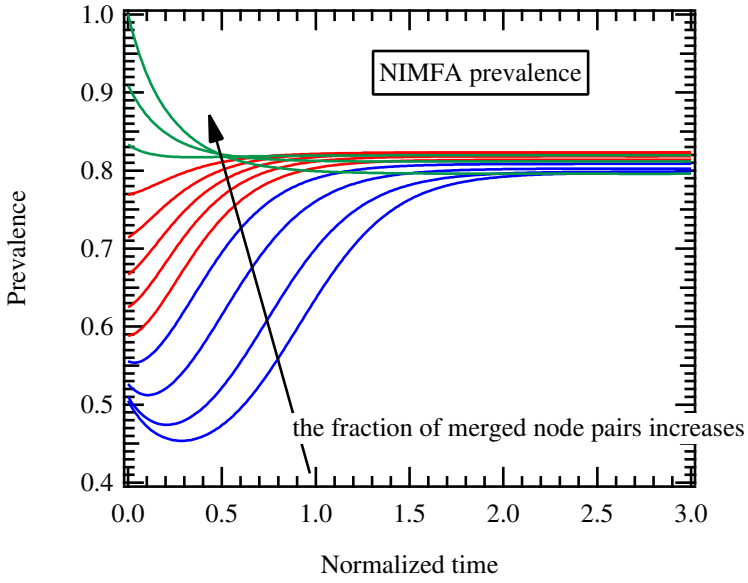

Figure 2.14: NIMFA prevalence corresponding to Fig. 2.13.

rapidly to the metastable state, which reflects the influence of the network structure on the spreading process.

2.5.2 Two cliques connected by links

Two cliques K_N can be connected by links. Each node in one clique is possible to be connected by a link to an arbitrary node in another clique. The weakest connection is that two complete graphs are connected by only one link, and the strongest connection is that all nodes are connected where the network is a complete graph K_{2N}. The fraction of connected node pairs between two cliques is denoted by b (i.e., if a node pair from each clique is arbitrarily selected, then the probability that the node pair connects is b). From (2.14), the prevalence decreases initially if

$$\tau < \frac{1}{bN} \tag{2.16}$$

If $b = 0$, then the whole network is disconnected and the prevalence decreases initially, and if $b = 1$, then the network is a complete graph K_{2N}.

As shown in Fig. 2.15, the Markovian SIS prevalence is non-unimodal when the initial cut size between the infected nodes and healthy nodes $|\partial G_I|$ is small. If the number of connecting links is below 500 (b=0.2) from (2.16), then the non-unimodal prevalence decreases initially and lately increases to the metastable state as the blue curves shown. However, different from the process on the network in Section 2.5.1, NIMFA performs badly to approximate the Markovian SIS prevalence when two cliques are connected with only a few links.

2.5.3 The barbell graph

In the two situations mentioned above, the connection between two cliques is fairly strong. The virus spread almost surely from one clique to another. A weaker connecting situation is that two complete graphs are connected by one path with length $L \geq 1$ and the network is a barbell graph. There is only one possible link which can contribute to the infection within a clique from another, and the SIS process within the clique is approximate a SIS process on a complete network.

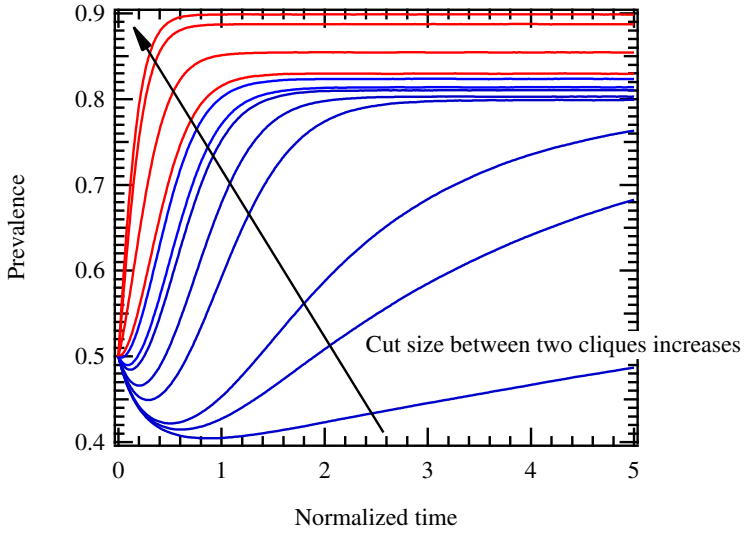

Figure 2.15: The prevalence of SIS process with $\tau = 0.1$ on networks which contain two cliques with size $N = 50$. The fraction of connecting links is $b = 1/2500, 1/500, 1/250, 1/50, 1/25, 2/25, 1/10, 4/25$ for blue curves and $b = 0.2, 0.4, 0.8, 1$ for red curves. Initially, all nodes in one clique are infected.

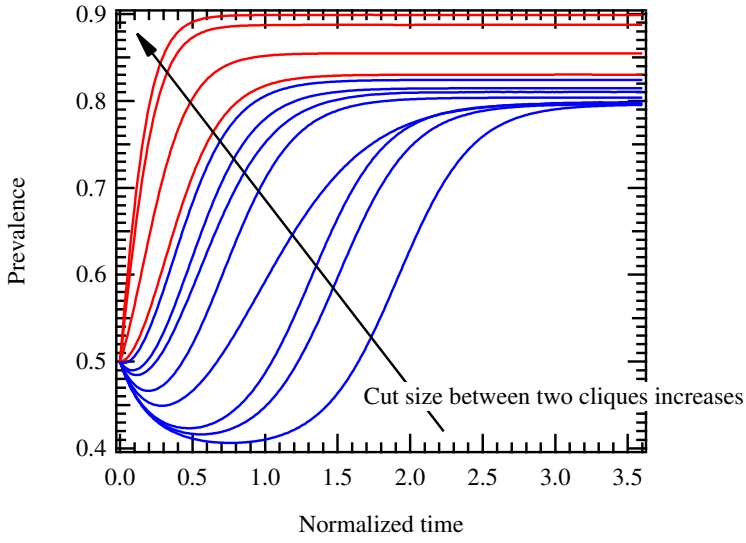

Figure 2.16: NIMFA prevalence corresponding to Fig. 2.15.

Initially, the virus spreads within one of the cliques. The virus may spread to another clique or die out for each realization of the SIS process, which depends on the relation between the extinction time of virus and the spreading time from one clique to another. The spreading time of the virus from one clique to another through an L-length path is not feasible to be obtained due to the exponentially large number of the nodal governing equations (2.2). However, we can obtain a lower bound of the mean spreading time by converting the SIS epidemic process on a path graph into a birth-death process. We index the nodes in an L-length (i.e., there are L links in the path) path from one end to another as $0, 1, 2, \cdots, L$. Firstly, we assume that the virus spreads from node L to node 0 and node L is infected without curing. A state of the birth-death process is the smallest index of the nodes which are infected. For example, if the process is in state k, then node k is infected and node 0 to node $k-1$ are all healthy. State 0 is an absorbing state where the virus successfully infects its destiny node 0. Figure 2.17 shows the birth-death process of the lower bound case. The death rate from state k to $k-1$ is the infection rate β, and the birth rate is the curing rate δ. The expected extinction time of the above-described birth-death process is a lower bound of the expected spreading time from node L to node 0 because of the following reason: the virus spreads backward only one step from node k to $k+1$ with rate δ for $k = 1, 2, \cdots, L-1$. In the SIS process on a path graph, the virus can spread backward more than one step from node k to an arbitrary node $k+1$ to L. As shown in Appendix A.1.2, the lower bound of the expected spreading time t_L through an L-length path is,

$$t_L = \frac{L}{\beta} + \frac{L\delta}{\beta(\beta-\delta)} - \frac{\delta\left(1-(\frac{\delta}{\beta})^L\right)}{(\beta-\delta)^2} \tag{2.17}$$

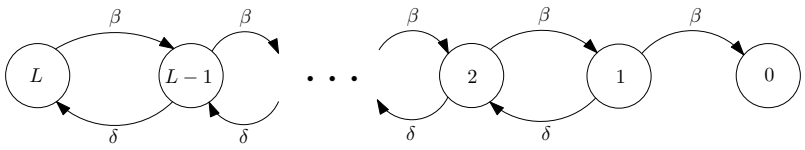

Figure 2.17: Schematic of the Markov chain for calculating the lower bound of spreading time over an L-length path.

We can also obtain an upper bound of the expected spreading time by a slight modification of the birth-death process. As shown in Fig. 2.18, we assume that the virus spreads back to the starting node L when the nearest infected node of node 0 is cured. In this Markov chain, the hitting time of the absorbing state is larger than the expected spreading time. The upper bound of

the expected spreading time through an L-length path is (see appendix A.1.2),

$$t_U = \frac{1}{\delta}\left(\left(\frac{\beta+\delta}{\beta}\right)^L - 1\right) \tag{2.18}$$

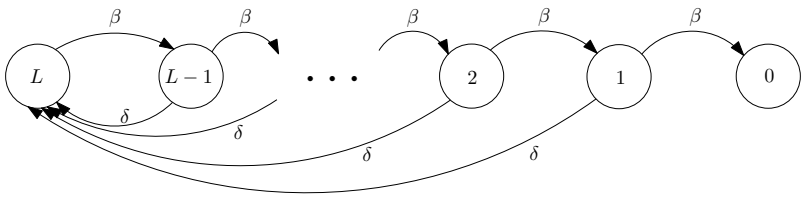

Figure 2.18: Schematic of the Markov chain for calculating the upper bound of spreading time over an L-length path.

Figure 2.19 and 2.20 show the lower bound and the upper bound of the expected spreading time with different path length L and effective infection rate τ. Figure 2.21 illustrates the expected extinction time of the SIS process on complete graphs. For a smaller (larger) effective infection rate τ, the expected spreading time through a path is larger (smaller) while the expected extinction time of the SIS process on a clique is smaller (larger). For two complete graphs which are connected by a path, the expected spreading time can be much larger than the extinction time, and then the virus cannot spread from one clique to another; the spreading time can also be much smaller that the virus can surely spread all over the network.

Figure 2.22 presents the prevalence of the Markovian SIS process and NIMFA on barbell graphs. The prevalence is non-unimodal when the connecting path length is small. For a long length path, the virus may never spread from one clique to another, and the virus spreading is restricted within only one clique. For NIMFA, the spreading time through paths linearly increases with the path length, and the NIMFA prevalence eventually converges to the globally asymptotic steady state [114]. Flat prevalence of NIMFA emerges with long paths. The Markovian SIS metastable state prevalence is at the same level of the metastable state NIMFA prevalence. The Markovian SIS process may never reach the level of the steady NIMFA prevalence, and the prevalence changes from a non-unimodal function to a unimodal one with the increase of the path length.

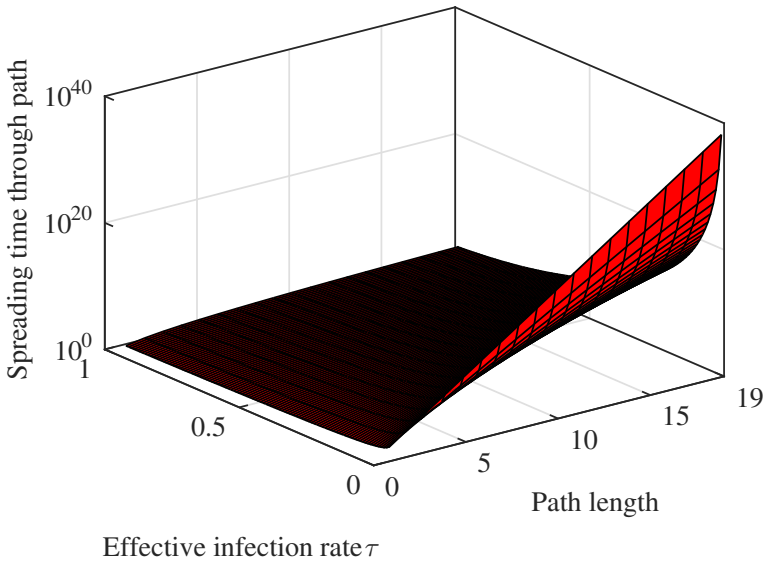

Figure 2.19: The lower bound of spreading time through a path with $\delta = 1$.

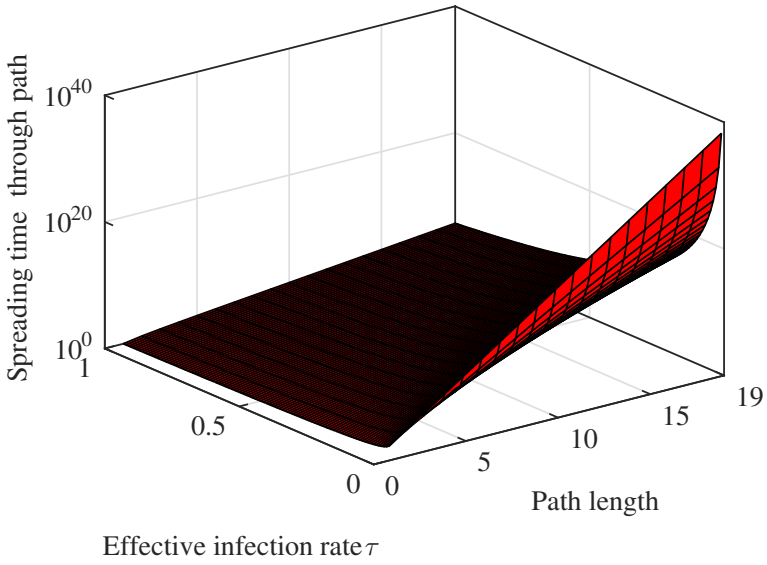

Figure 2.20: The upper bound of spreading time through a path with $\delta = 1$.

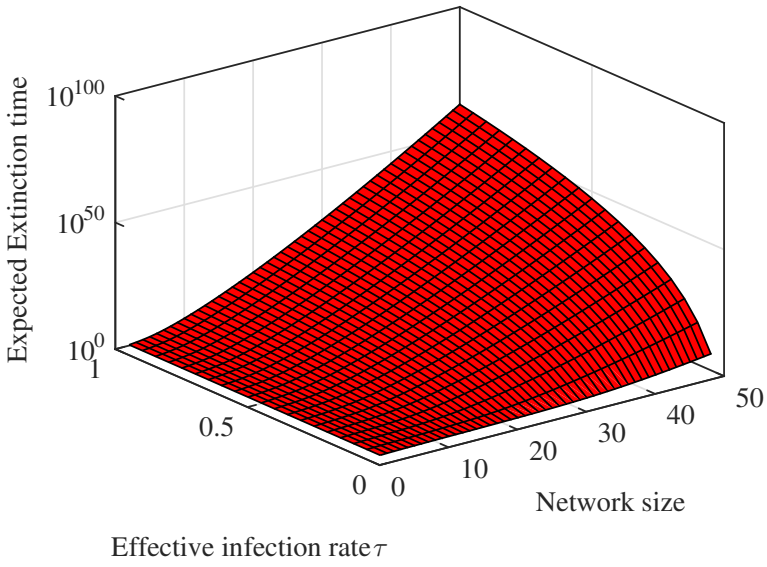

Figure 2.21: The expected extinction time of SIS process on complete graphs.

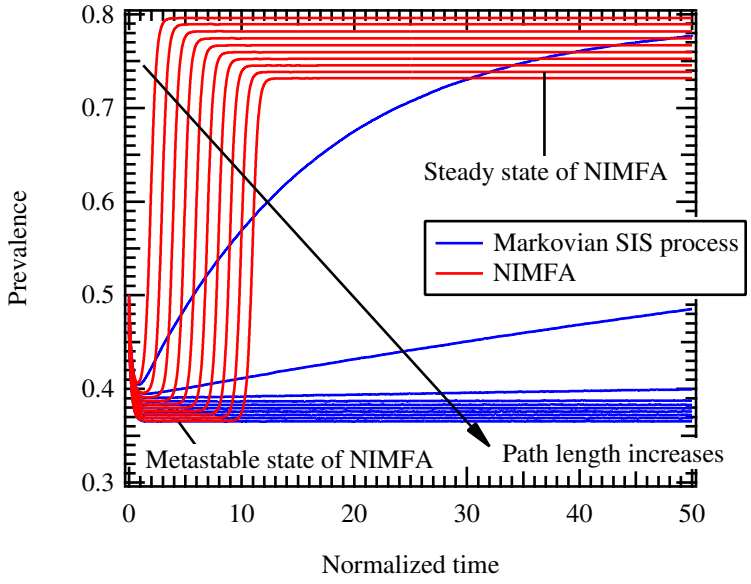

Figure 2.22: Prevalence of Markovian SIS process and NIMFA on barbell graphs with path length from 1 to 10 and effective infection rate $\tau = 0.1$.

2.5.4 More cliques

By increasing the number of cliques, multiple flat stages of NIMFA prevalence appear. Figure 2.23 shows the prevalence of the SIS process with $\tau = 0.05, 0.1, 1$ on a network whose 6 cliques with size 50 are connected by 5 paths with length 5. There are 5 flat phases and one steady phase. The Markovian SIS prevalence stays at the same low level of the NIMFA prevalence of the first flat phase when τ is low, i.e., $\tau = 0.05, 0.1$. If τ is large, then the prevalence increases to the metastable state almost linearly with time t, which is much slower than the NIMFA prevalence.

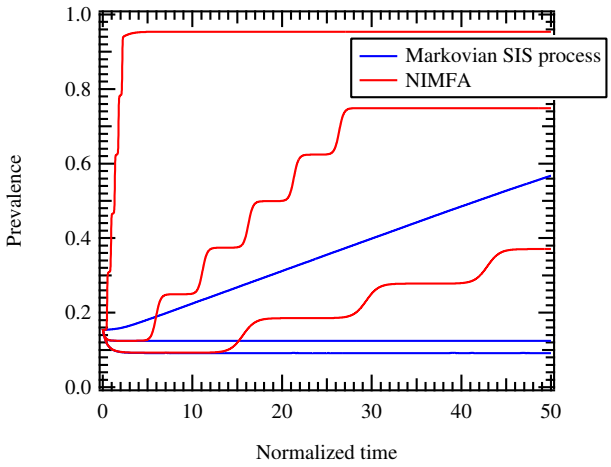

Figure 2.23: The SIS process with infection rate $\tau = 0.05, 0.1, 1$ on a network with 6 cliques and each clique with size 50 is connected one by one with a path of length 5. Initially all nodes in one of the cliques are infected.

2.6 Conclusion

In this chapter, we reviewed several complex network models in a brief way and introduced the networked SIS spreading model. The mean-field analysis and the simulation method of the SIS model are discussed. Our further studies are based on those backgrounds. Moreover, the inaccuracy of NIMFA due to the early die-out, which is sometimes neglected in simulations, is pointed out. Finally, we show that: different from homogeneously mixed SIS models, the time-dependent prevalence can be non-unimodal due to the underlying networks.

3

Autocorrelation of the SIS Process

In this chapter, we focus on the autocorrelation of the Susceptible-Infected-Susceptible (SIS) process on networks. NIMFA is applied to calculate the autocorrelation properties of the exact SIS process. We derive the autocorrelation of the infection state of each node and the fraction of infected nodes both in the steady and transient state, as functions of the infection probabilities of nodes. Moreover, we show that the autocorrelation can be used to estimate the infection and curing rates of the SIS process. The theoretical results are compared with the simulation of the exact SIS process. Our work fully utilizes the potential of the mean-field method in a novel way and shows that NIMFA can indeed capture the autocorrelation properties of the exact SIS process.[1]

3.1 Mean-field approach to autocorrelation

LOCALLY, an individual node in the network can be infected and cured repeatedly so that the infection state $X_j(t)$ at two different time points can be autocorrelated. The autocorrelation of the infection state of a node j between time s and t is

$$\rho_j(s,t) \triangleq \frac{E[X_j(s)X_j(t)] - E[X_j(s)]E[X_j(t)]}{\sqrt{\mathrm{Var}[X_j(s)]\mathrm{Var}[X_j(t)]}} \tag{3.1}$$

The numerator on the right-hand side in (3.1) is the covariance of the infection state $X_j(s)$ and $X_j(t)$, and the denominator normalizes the covariance. If time $t = s$, then the infection state $X_j(s)$ is fully correlated with itself, and the autocorrelation is $\rho_j(s,s) = 1$. If $X_j(s)$ and $X_j(t)$ are independent, then the autocorrelation is $\rho_j(s,t) = 0$. The autocorrelation is symmetric: $\rho_j(s,t) = \rho_j(t,s)$. Given the initial infection state of the network

[1] This chapter is based on [24].

$[X_1(0), \ldots, X_N(0)]$, the infection states $X_j(s)$ and $X_j(t)$ are positively correlated [113, Corollary 1] such that $E[X_j(s)X_j(t)] \geq E[X_j(s)]E[X_j(t)]$ and the autocorrelation $\rho_j(s,t) \geq 0$. The autocorrelation $\rho_j(s,t)$ contains one second-moment term $E[X_j(s)X_j(t)]$, but the rest of the terms can be calculated given the first-moment infection probabilities $E[X_j(s)]$ and $E[X_j(t)]$.

The autocorrelation contains information about the change of the infection state of each node. A large autocorrelation implies that the change of the infection state is slow, and the infection state is more likely to be identical between time s and t. While a smaller autocorrelation indicates that the infection state between time s and t is more independent. Globally, the fluctuating fraction of infected nodes $I(t) \triangleq \frac{1}{N}\sum_{j=1}^{N} X_j(t)$ is also autocorrelated, and the autocorrelation and its spectral analysis of $I(t)$ in real epidemics can be traced back to Anderson et al. [143]. By analyzing the autocorrelation and its spectrum of the incidence data of pertussis, mumps, and measles, Anderson et al. [143] indicate statistically significant seasonal and the longer-term resurgence of those diseases and find that vaccination increases the periods of the longer-term oscillations of the incidence data. However, in the basic networked SIS model, the autocorrelation of the infection state is infeasible to be calculated, because the SIS model is a 2^N-state Markov process [102, 103] and the computational complexity is exponentially high regarding of the network size N. Previously, Meier et al. [144, Supplementary Information E] analyzed the correlation of the infection state of the SIS model for small time intervals, but the calculation involves higher-order moments. In this chapter, we apply NIMFA to study the autocorrelation of the infection state $X_j(t)$ and the fraction of infected nodes $I(t)$ both in the transient and steady states. Particularly in the steady state, we derive the explicit formula of the autocorrelation of the infection state, which is an exponentially decreasing function of time delay. The accuracy of the NIMFA autocorrelation is evaluated by simulating the exact SIS process. The result indicates that NIMFA, as an approximate stochastic process, well captures the autocorrelation properties of the exact SIS process. Moreover, the autocorrelation can also be used to estimate the infection and curing rates of the SIS process.

3.2 Autocorrelation in the steady state

In the steady state, the NIMFA autocorrelation of the infection state of node j with time lag h is defined by

$$R_{j\infty}(h) \triangleq \frac{E\left[V_{j\infty}(t)V_{j\infty}(t+h)\right] - v_{j\infty}^2}{\mathrm{Var}[V_{j\infty}(t)]} \tag{3.2}$$

where $\mathrm{Var}[V_{j\infty}(t)] = v_{j\infty} - v_{j\infty}^2$ since $V_{j\infty}(t) \in \{0, 1\}$ is a Bernoulli random variable. By further derivation (see Appendix A.2.1), we obtain the autocorrelation as a function of the steady-state infection probability $v_{j\infty}$ and the curing rate δ,

$$R_{j\infty}(h) = \mathrm{e}^{-\frac{\delta}{1-v_{j\infty}}h} \tag{3.3}$$

where we assume that the time lag h is positive without loss of generality. Since the autocorrelation is symmetric $R_{j\infty}(h) = R_{j\infty}(-h)$, $R_{j\infty}(h) = \mathrm{e}^{\frac{\delta}{1-v_{j\infty}}h}$ for $h < 0$. The NIMFA infection probability $v_{j\infty}$ in (3.3) can be obtained by solving the NIMFA Eq. (2.4) numerically.

With a fixed δ and time lag h, the autocorrelation $R_{j\infty}(h)$ in (3.3) decreases with the infection rate β because the infection probability $v_{j\infty}$ increases correspondingly. A larger infection rate β implies a faster state transition from healthy to infected, and the autocorrelation of the infection state is smaller consequently. A larger δ leads to a faster transition from infected to healthy, but, simultaneously, the infection probability of each neighbor becomes smaller. Therefore, the state transition of each node is slower from the healthy state to the infected state, and the effect of the curing rate δ is unclear. Only in special networks, the effect of the curing rate δ can be determined. For example, the infection probabilities of all nodes are equal to $1 - \frac{1}{k\tau}$ in a k-regular graph [102], and then the autocorrelation function becomes

$$R_{j\infty;k\text{-regular}}(h) = \mathrm{e}^{-\beta k h} \tag{3.4}$$

Formula (3.4) indicates that the autocorrelation of the infection state does not depend on the curing rate δ in regular graphs, which enables to adjust the autocorrelation while keeping the effective infection rate τ unchanged. In regular graphs, the effect of the decrease (increase) of $v_{j\infty}$ is exactly compensated by the increase (decrease) of δ in (3.3). The autocorrelation under other mean-field approximations can also be derived with the same procedure. For example, the Heterogeneous Mean-Field approximation (HMF) assumes statistical equivalence among the nodes with the same degree [12], and the

autocorrelation under HMF has the same form as the NIMFA autocorrelation (see Appendix A.2.2). In the case of regular graphs, HMF is equivalent [145] to NIMFA and then their approximate autocorrelations are identical.

Generally, the NIMFA infection probability of node j with degree d_j for $j = 1, \cdots, N$ is bounded by [102]

$$1 - \frac{1}{1 + \tau d_j - \frac{d_j}{d_{\min}}} \leq v_{j\infty} \leq 1 - \frac{1}{1 + \tau d_j}$$

in a connected network with minimum degree $d_{\min}$, and the NIMFA autocorrelation (3.3) is thus bounded by

$$\mathrm{e}^{-(1+\tau d_j)\delta h} \leq R_{j\infty}(h) \leq \mathrm{e}^{-\left(1+d_j\left(\tau-\frac{1}{d_{\min}}\right)\right)\delta h} \tag{3.5}$$

The largest eigenvalue of the adjacency matrix λ_1 follows $\lambda_1 \geq d_{\min}$, and then the effective infection rate τ can either be larger or smaller than $1/d_{\min}$ when τ is above the threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$. Equation (3.3) indicates that the autocorrelation has another upper bound

$$R_{j\infty} < \mathrm{e}^{-\delta h} \tag{3.6}$$

when $v_{j\infty} > 0$ (i.e. above the threshold). If $\frac{1}{\lambda_1} < \tau < \frac{1}{d_{\min}}$, then

$$\mathrm{e}^{-\delta h} < \mathrm{e}^{-\left(1+d_j\left(\tau-\frac{1}{d_{\min}}\right)\right)\delta h}$$

and the upper bound (3.6) is tighter. If $\tau > \frac{1}{d_{\min}}$, then the upper bound in (3.5) is tighter, and we can rewrite (3.5) as

$$\mathrm{e}^{-(1+\tau d_j)\delta h} \leq R_{j\infty}(h) \leq \mathrm{e}^{-(1+\tau d_j)\delta h} \cdot \mathrm{e}^{\left(\frac{d_j}{d_{\min}}\right)\delta h} \tag{3.7}$$

In (3.7), the upper bound is just the product of the lower bound and the term $\mathrm{e}^{(d_j/d_{\min})\delta h} > 1$. In a network with large degree deviation $d_j/d_{\min}$, the bound (3.7) is loose. In the regular graph, $\lambda_1 = d_{\min}$, and the upper bound achieves the exact NIMFA autocorrelation (3.4) while the lower bound does not.

In a heterogeneous network, e.g. the scale-free network, the degree d_j can diverge in the thermodynamic limit $N \to \infty$. Thus, if $\tau > \frac{1}{d_{\min}}$ and $d_j \to \infty$, then both the upper and lower bound in (3.7) converge to zero, and the autocorrelation $R_{j\infty}(h) = 0$. If $\frac{1}{\lambda_1} < \tau < \frac{1}{d_{\min}}$ and $d_j \to \infty$, then the lower bound $R_{j\infty}(h) > \mathrm{e}^{-(1+\tau d_j)\delta h}$ converges to zero. Consequently, the autocorrelation is loosely bounded by $0 \leq R_{j\infty}(h) \leq \exp(-\delta h)$.

From a global point of view, the fraction of infected nodes $I(t) = \frac{1}{N}\sum_{j=1}^{N} X_j(t)$ in the steady state can be approximated by $I_\infty^{(1)}(t) \triangleq \frac{1}{N}\sum_{j=1}^{N} V_{j\infty}(t)$. The autocorrelation of $I_\infty^{(1)}(t)$ is just a linear combination of the autocorrelation of each node (see Appendix A.2.1),

$$R_{I^{(1)}\infty}(h) = \frac{\sum_{j=1}^{N}\left(v_{j\infty} - v_{j\infty}^2\right) R_{j\infty}(h)}{\sum_{j=1}^{N}(v_{j\infty} - v_{j\infty}^2)} \tag{3.8}$$

3.3 Autocorrelation in the transient state

In this section, we consider the NIMFA autocorrelation of the SIS process at two arbitrary time points s and t, respectively. Different from that in the steady state in Section 3.2, the infinitesimal generator (2.6) is a determined function of time given the initial state. The 2-state Markov process (2.5) of each node is thus a time-inhomogeneous process. Calculating the process (2.5) allows us to analyze the autocorrelation of the epidemic process in the transient regime before the metastable state, or the regime before the all-healthy steady state when the effective infection rate $\tau < \tau_c^{(1)}$.

We denote the NIMFA autocorrelation of node j between time s and t by

$$R_j(s,t) \triangleq \frac{E[V_j(s)V_j(t)] - v_j(s)v_j(t)}{\sqrt{[v_j(s) - v_j^2(s)][v_j(t) - v_j^2(t)]}} \tag{3.9}$$

Following a similar derivation as Eq. (3.8) in the steady state, the autocorrelation of the fraction of infected nodes $R_{I^{(1)}}(s,t)$ is also a linear combination of the autocorrelation of each node,

$$R_{I^{(1)}}(s,t) = \frac{\sum_{j=1}^{N}\sqrt{[v_j(s) - v_j^2(s)][v_j(t) - v_j^2(t)]}R_j(s,t)}{\sqrt{\sum_{j=1}^{N}[v_j(s) - v_j^2(s)]\sum_{j=1}^{N}[v_j(t) - v_j^2(t)]}} \tag{3.10}$$

Similar to the steady-state autocorrelation in Section 3.2, we only use the infection probabilities in the calculation, and the joint expectation $E[V_j(s)V_j(t)]$ in (3.9) becomes a crucial term. The calculation of the joint expectation $E[V_j(s)V_j(t)] = v_j(s)\Pr[V_j(t) = 1|V_j(s) = 1]$ involves the 2×2 time-dependent transition matrix $P_j(s,t)$ of which the element is $(P_j(s,t))_{kl} = \Pr[V_j(t) = l - 1|V_j(s) = k - 1]$. The computation of the

autocorrelation functions (3.9) and (3.10), requires us first to calculate the matrix $P_j(s,t)$.

The matrix $P_j(s,t)$ follows the time-inhomogeneous Kolmogorov forward equation

$$\frac{dP_j(s,t)}{dt} = P_j(s,t)Q_j(t) \tag{3.11}$$

where $Q_j(t)$ is the NIMFA infinitesimal generator (2.6). We can apply the Magnus expansion [146, 147] to analyze the NIMFA transition matrix $P_j(s,t)$ in Eq. (3.11). A brief introduction of the Magnus expansion can be found in Appendix A.2.3. Although the calculation of the exact NIMFA transition probability $P_j(s,t)$ is not possible, approximations of $P_j(s,t)$ allowing a fair comparison between NIMFA and the exact SIS process can be made with restricted error. First, there exists a 2×2 matrix $\Omega(s,t;j)$ such that the solution of Eq. (3.11) is $P_j(s,t) = \exp(\Omega(s,t;j))$. Second, if (see the derivation of (A.18) for details in Appendix A.2.3)

$$0 < t - s < T \triangleq \frac{\pi}{\sqrt{\beta^2 d_j^2 + \delta^2}} \tag{3.12}$$

then the exponent matrix $\Omega(s,t;j)$ can be expanded into a convergent Magnus series $\Omega(s,t;j) = \sum_{k=1}^{\infty} \Omega_k(s,t;j)$. Specifically, by only preserving the first term, i.e. $\Omega_1(s,t;j) = \int_s^{s+h} Q_j(t)dt$, in the convergent Magnus series of $\Omega(s,t;j)$, we can achieve a third-order accuracy (see Appendix A.2.3) for the time length $h = t - s$, i.e.

$$P_j(s,s+h) \;=\; \exp\left(\int_s^{s+h} Q_j(t)dt\right) + O(h^3) \tag{3.13}$$

Equation (3.13) holds because $\exp(X+O(h^k)) = \exp(X)+O(h^k)$ holds for a matrix X as can be verified by evaluating their power series. Using the Taylor expansion of the infinitesimal generator $Q_j(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \left.\frac{d^k Q_k(u)}{du^k}\right|_{u=s} (t-s)^k$ at time s, the solution (3.13) becomes

$$P_j(s,s+h) = \exp\left(Q_j(s)h + \left.\frac{dQ_j(t)}{dt}\right|_{t=s} \frac{h^2}{2}\right) + O(h^3) \tag{3.14}$$

Only the first two terms of the Taylor expansion of the infinitesimal generator $Q_j(t)$ are preserved in (3.14) since the error is $O(h^3)$ in (3.13). The first term on the right-hand side of (3.14) can be calculated by matrix diagonalization

described in Appendix A.2.1. The derivative of the infinitesimal generator $dQ_j(t)/dt$ involves $\frac{d\tilde{\beta}_j(t)}{dt}$ from Eq. (2.6), which is

$$\begin{aligned}\tilde{\beta}'_j(t) &\triangleq \frac{d\tilde{\beta}_j(t)}{dt} \\ &= \beta \sum_{i\in\mathcal{N}_j} \frac{dv_i(t)}{dt} \\ &= \beta \sum_{i\in\mathcal{N}_j} \left(-\delta v_i(t) + \beta[1-v_i(t)] \sum_{k\in\mathcal{N}_i} v_k(t) \right)\end{aligned} \tag{3.15}$$

where $\mathcal{N}_j$ denotes the neighbors of node j. The calculation in Eq. (3.15) involves the infection probabilities of 2-hop neighbors of node j. Specifically, the transition probability that node j remains infected after h time units is

$$\begin{aligned}&\Pr[V_j(s+h)=1|V_j(s)=1] \\ =&(P_j(s,s+h))_{22} \\ =&\frac{2\delta \mathrm{e}^{-(\tilde{\beta}_j(s)+\delta)h-\tilde{\beta}'_j(s)h^2/2} + 2\tilde{\beta}_j(s) + \tilde{\beta}'_j(s)h}{2\tilde{\beta}_j(s)+2\delta+\tilde{\beta}'_j(s)h} + O(h^3)\end{aligned}$$

Different from that in the steady state (see Eq. (3.3)), the infection probabilities of neighbors of node j always appear in the calculation of the transition matrix $P_j(s,t)$ in the transient state as indicated in (3.15). Higher-order accuracy is also possible by preserving more terms of the Magnus series, and higher-order derivative $d^k\tilde{\beta}_j(t)/dt$, which can be calculated by the infection probabilities of all nodes within $k+1$ hops from node j, is involved. For example, if we preserve the second term in the Magnus expansion of $\Omega(s,s+h;j)$, which can be calculated by the Taylor expansion as,

$$\begin{aligned}&\Omega_2(s,s+h;j) \\ =&\delta \begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix} \int_s^{s+h} dt_1 \int_s^{t_1} dt_2 \left(\tilde{\beta}_j(t_1) - \tilde{\beta}_j(t_2)\right) \\ =&\delta \begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix} \left(\frac{1}{6}\frac{d\tilde{\beta}_j(t)}{dt}\bigg|_{t=s} h^3 + \frac{1}{12}\frac{d^2\tilde{\beta}_j(t)}{dt^2}\bigg|_{t=s} h^4 \right) + O(h^5)\end{aligned}$$

then we can achieve an accuracy of $O(h^5)$ because (see Appendix A.2.3)

$$P_j(s,s+h) = \exp\left(\sum_{i=1}^{2} \Omega_i(s,s+h;j)\right) + O(h^5)$$

and the calculation involves the infection probabilities of neighbors within 3 hops. For NIMFA, preserving more terms is not always reasonable, because the infection probability of each node can only be solved numerically. When more Magnus terms are preserved, the inaccuracy is mainly caused by the numerical method which solves the nonlinear NIMFA Eq. (2.4). For example, using the fourth-order Runge-Kutta method [147, p. 200], the error of the infection probabilities is of order $O(h^4)$.

For a time interval $t - s > T$, the Magnus expansion of the exponent $\Omega(s, t; j)$ may not converge. The time interval (s, t) can be divided into subintervals with length $h < T$ in which the Magnus series converges. The NIMFA transition matrix between time s and t can be written as

$$P_j(s,t) = \prod_{k=1}^{(t-s)/h} P_j(s + (k-1)h, s + kh) \tag{3.16}$$

by the Chapman-Kolmogorov Equation (see Eq. (A.15)). Equation (3.16) is also applicable to a small time interval $t - s < T$ to obtain a more accurate result. A r-th order accuracy regarding the time delay h is achieved for the transition matrix $P_j(s, t)$ using Eq. (3.16) if the accuracy is $O(h^{r+1})$ for each $P_j(s + (k-1)h, s + kh)$.

The analysis in this section allows us to calculate and compare the NIMFA autocorrelation with the exact SIS process since the error can be controlled, even though the exact NIMFA autocorrelation is not feasible in the transient state.

3.4 Numerical and simulation results

In this section, we compare the NIMFA autocorrelation with the autocorrelation of the exact SIS process from the simulation. The simulation of the exact SIS process is implemented by the Gillespie algorithm (Monte Carlo method) [25, 136, 148] and the theoretical results are obtained by solving the NIMFA Eq. (2.4) numerically (fourth-order Runge-Kutta method [147, p. 200]). In the steady state, we run the simulation for 40000 time units with the curing rate $\delta = 1$ and sample the infection state of each node every 0.001 time unit. In other words, we obtain the infection state $X_j(n/1000)$ for $n = 0, 1, \ldots, 4 \times 10^7$ from simulation. We only use the state sequence sampled after $t = 10000$ to ensure that the SIS process is in the metastable state. Moreover, the time series of the fraction of infected nodes can be calculated as $I(n/1000) = \frac{1}{N}\sum_{j=1}^{N} X_j(n/1000)$. In the transient state, 10^4

realizations of the infection states $X_j(s)$ and $X_j(t)$ are obtained to calculate the autocorrelation between two arbitrary time s and t.

3.4.1 Steady state

Figure 3.1 to 3.3 show the NIMFA autocorrelation and the simulated autocorrelation of the infection state of randomly selected nodes in an Erdős-Rényi (ER) graph, a regular graph with degree 26, and a star graph, respectively. The NIMFA autocorrelation $R_{j\infty}(h)$ is a very accurate approximation on those graphs. Figure 3.1 shows that the autocorrelation of the infection state is not sensitive to the value of the curing rate δ, which is reasonable because the deviation of the degree is small and the result is similar to that of the regular graph in Fig. 3.2. In Fig. 3.2, the autocorrelation of the infection state is identical to formula (3.4) that the autocorrelation is invariant to the curing rate δ in regular graphs. Figure 3.3 shows the autocorrelation of the infection state in a star graph. The autocorrelation of the hub node is much smaller than that of the leaf nodes since the infection probability of the hub node is larger. The cross-correlation of the infection states between neighbors shown in Fig. 3.1 to 3.3 is approximately 0, which leads to the effectiveness of NIMFA since NIMFA omits the cross-correlation between neighbors.

Figure 3.4 shows the autocorrelation of the infection state of a node in a cycle graph and NIMFA fails to capture the autocorrelation. Actually, NIMFA also fails to approximate the prevalence as shown in Fig. 3.4. In the situation of the cycle graph, the cross-correlation of the infection states between neighbors is much larger than zero and NIMFA itself is a bad approximation. The accuracy of Mean-Field methods has been studied in [149, 150, 151], which is beyond the scope of this chapter.

We also calculate the autocorrelation of the fraction of infected nodes $R_{I^{(1)}}(h)$. Figure 3.5 shows that NIMFA can also approximate the autocorrelation of the fraction of infected nodes in the star graph corresponding to Fig. 3.3.

3.4.2 Transient state

In the transient state, we validate the NIMFA autocorrelation on the star graph where the NIMFA infection probabilities are accurate while nodes have very different degrees.

Figure 3.6 shows the joint expectation of the infection states

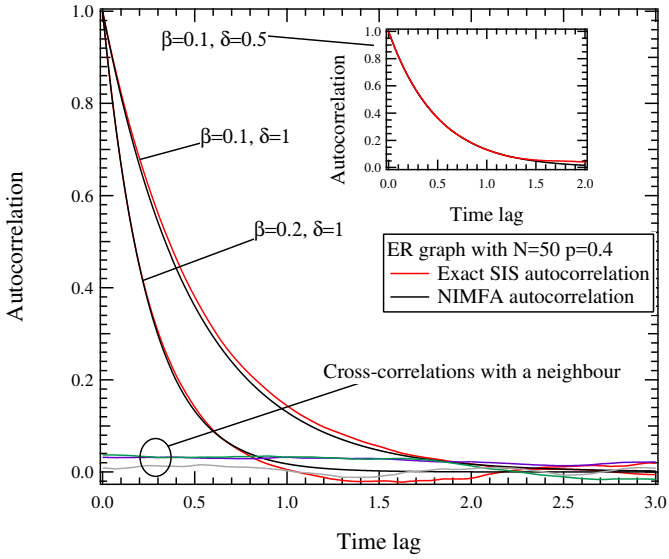

Figure 3.1: A randomly selected node is evaluated in an Erdős-Rényi (ER) network with the link connecting probability 0.4 and $N = 50$. The autocorrelation is approximately constant for different value of δ. The cross-correlation between the node with one of its neighbour is also plotted, which is almost zero.

$E[X_j(0)X_j(h)]$ and the corresponding NIMFA approximation $E[V_j(0)V_j(h)]$ of the leaf and hub nodes. For the leaf node and the hub node, the convergent time delay h of the Magnus series of $\Omega(s, s+h)$ are $h < T \approx 2.221$ and $h < T \approx 0.064$ from (3.12), respectively. Figure 3.6 indicates that the NIMFA joint expectation $E[V_j(0)V_j(h)]$ (the blue lower curve) is accurate comparing with the exact joint expectation $E[X_j(0)X_j(h)]$ for a small time delay h, i.e. $h < 0.2$ for the leaf node. For a large time delay, the inaccuracy is due to either the omission of term $O(h^3)$ in (3.14) or that the NIMFA transition probability matrix $P_j(s,t)$ itself is a bad approximation, but we can eliminate the possibility of the latter using Eq. (3.16). As the black middle curve in Fig. 3.6 indicated, the NIMFA joint expectation $E[V_j(0)V_j(h)]$ is indeed a good approximation using Eq. (3.16) with subinterval length 0.01.

From a global point of view of the network, Fig. 3.7 presents the autocorrelation of the fraction of infected nodes $R_I(0.5, 0.5+h)$ and the corresponding NIMFA approximation $R_{I^{(1)}}(0.5, 0.5+h)$, which are in the transient state of the SIS process before the metastable state. The exact autocorrelation is well fitted by NIMFA. Interestingly, the decay of the autocorrelation in the transient state is also exponential as shown in Fig. 3.7, but we cannot demonstrate

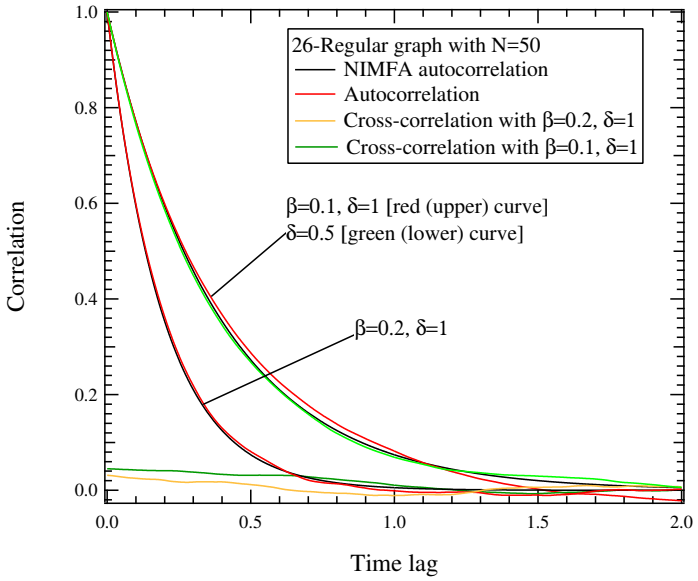

Figure 3.2: The autocorrelation of the infection state of a 26-regular graph with $N = 50$. The results are similar to those of the ER graph. The autocorrelation is invariant to δ.

exponential decay as opposed to the steady state.

In this section, we have tested our method on different networks with size 50, but for larger networks, the results are similar. In a conclusion, NIMFA captures the autocorrelation properties of the exact SIS process except in the cases that NIMFA is not applicable even for approximating the first-moment properties, i.e. the infection probabilities $E[X_j(t)]$ and the prevalence $y(t)$.

3.5 Estimating the curing and infection rates

In real epidemics, a disease agency may have the infection-state data by monitoring individuals periodically but no information about the rates. We consider the reverse problem of estimating the curing rate δ and the infection rate β, given the sequence $X_j(t+\Delta), X_j(t+2\Delta), \ldots, X_j(t+n\Delta)$ of the infection state of node j in the metastable state. From Eq. (3.3), the curing rate is

$$\delta = -(1 - v_{j\infty}) \frac{\ln(R_{j\infty}(h))}{h} \tag{3.17}$$

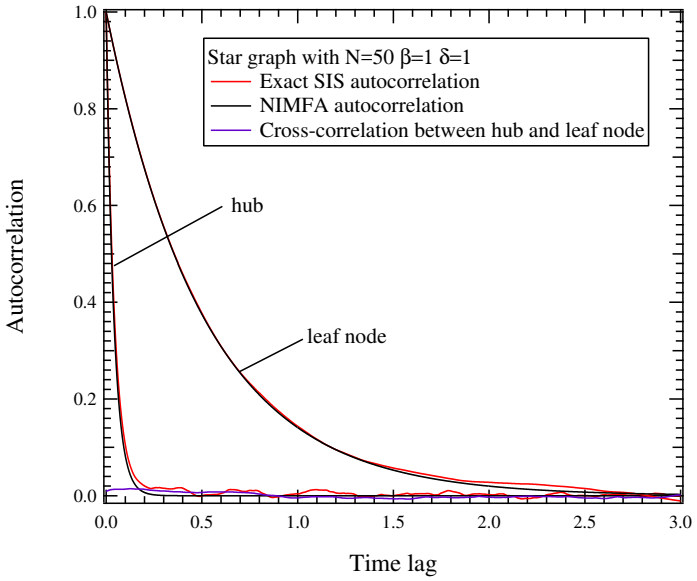

Figure 3.3: The autocorrelation of the infection state of the hub and a leaf node in a star graph with $N = 50$. The NIMFA autocorrelation shows a very good approximation and the cross-correlation between hub and leaf nodes is approximate 0.

Formula (3.17) can be used to estimate the curing rate δ of the SIS process. In formula (3.17), we can approximate the infection probability as $v_{j\infty} \approx \frac{1}{n}\sum_{i=1}^{n} E[X_j(t+i\Delta)]$, while the autocorrelation $R_{j\infty}(h)$, which approximates the exact autocorrelation $\rho_j(s,t)$ in (3.1), is just the autocorrelation of the binary infection sequence $X_j(t+k\Delta)$. Furthermore, using the NIMFA equation in the metastable state $-\delta v_{j\infty} + \beta(1-v_{j\infty})\sum_{i=1}^{N} a_{ji}v_{i\infty} = 0$, we can eliminate δ and (3.17) becomes

$$\beta = -\frac{v_{j\infty}}{\sum_{i=1}^{N} a_{ji}v_{i\infty}} \frac{\ln(R_{j\infty}(h))}{h} \tag{3.18}$$

Under NIMFA, the curing rate δ can be estimated by (3.17) without knowing the underlying network. However, to estimate the infection rate β, formula (3.18) involves the network information. We rewrite (3.18) as

$$-\frac{v_{j\infty}}{\beta}\frac{\ln(R_{j\infty}(h))}{h} = \sum_{i=1}^{N} a_{ji}v_{i\infty}$$

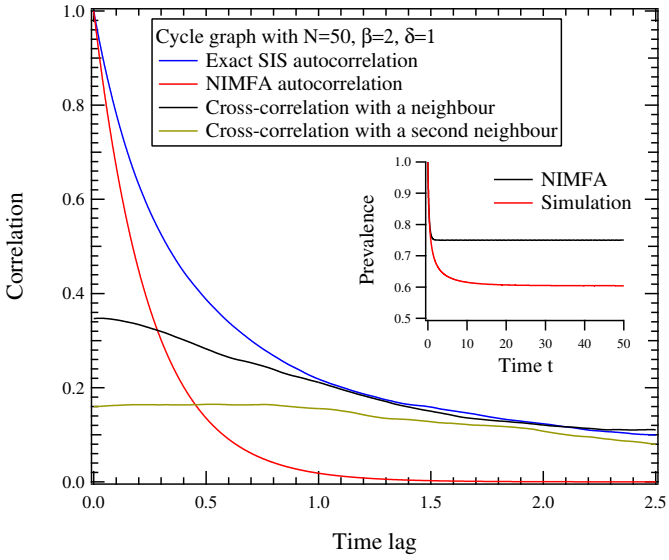

Figure 3.4: The correlation of the infection state of a node and the prevalence in a cycle graph with $N = 50$. Initially all nodes are infected to prevent the inaccuracy caused by early die-out [26]. The NIMFA autocorrelation is much smaller than the exact one, and the cross-correlations between neighbours and second-hop neighbours are very large.

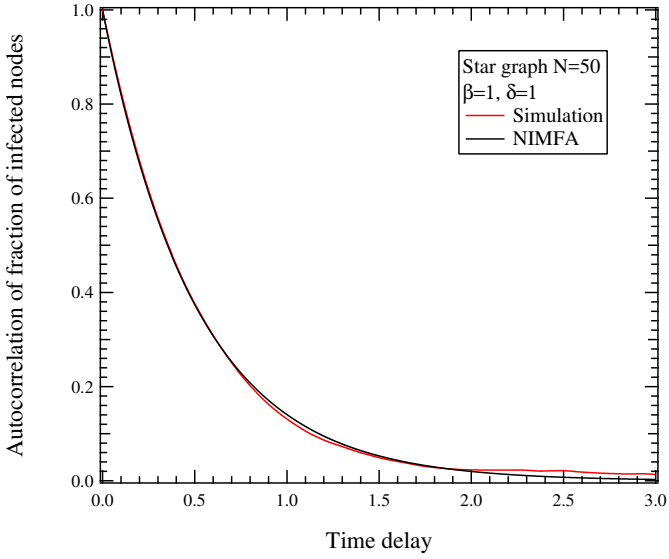

Figure 3.5: The autocorrelation of the fraction of infected nodes $I(t)$ in the metastable state.

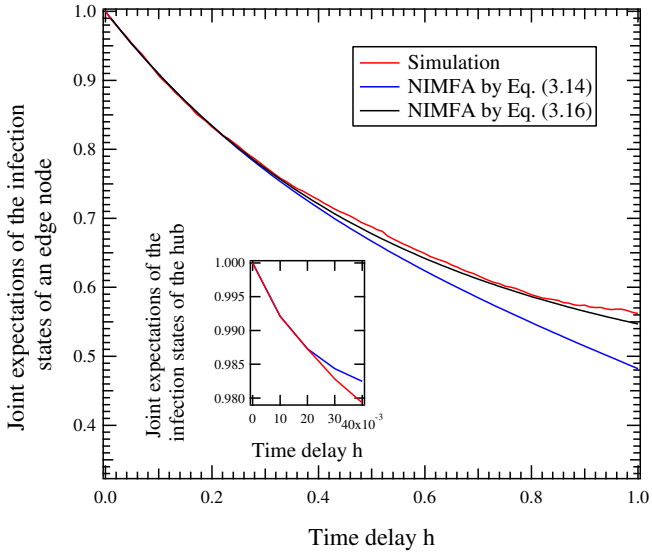

Figure 3.6: The joint expectation of the infection state $E[X_j(0)X_j(h)]$ and the corresponding NIMFA approximation $E[V_j(0)V_j(h)]$ of the SIS process on the star graph.

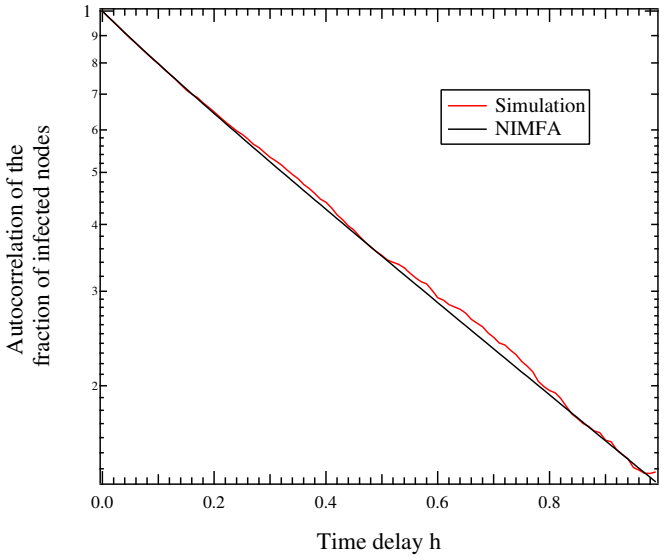

Figure 3.7: The autocorrelation of the fraction of infected nodes $R_I(0.5, 0.5+h)$ and the corresponding NIMFA approximation $R_{I^{(1)}}(0.5, 0.5+h)$ of the SIS process on the star graph.

and sum over all nodes

$$-\sum_{j=1}^{N}\frac{v_{j\infty}}{\beta}\frac{\ln(R_{j\infty}(h))}{h}=\sum_{j=1}^{N}\sum_{i=1}^{N}a_{ji}v_{i\infty}=\sum_{i=1}^{N}d_iv_{i\infty}$$

where d_i is the degree of node i. After rearrangement of the above equation, we obtain,

$$\beta=-\frac{\sum_{j=1}^{N}v_{j\infty}}{\sum_{j=1}^{N}d_jv_{j\infty}}\frac{\ln(R_{j\infty}(h))}{h} \tag{3.19}$$

Thus, the estimation of the infection rate β requires either the degree of every node d_j for all j as in (3.19) or the local topology information about node j, i.e. a_{ji} for all i as in (3.18).

Using the binary infection-state sequence $X_j(t+k\Delta)$ obtained by simulation, we estimate the curing rate δ and the infection rate β by (3.17) and (3.18), respectively. In Fig. 3.8, the value of the estimated rates times the time lag h is plotted for a leaf node of the star graph corresponding to Fig. 3.3. The slopes of the linear fitting functions (red curves in Fig. 3.8) are the estimated rates, and both the estimated infection rate β and the curing rate δ are 1.00 while both the real rates equal to 1.

3.6 Conclusion

In this chapter, we study the autocorrelation, the only second-moment property captured by NIMFA, of the SIS process. We obtained the explicit formula of the autocorrelation, i.e. Eq. (3.3), under NIMFA in the steady state, and the steady-state autocorrelation follows an exponential decay with the time lag. Interestingly, the steady-state autocorrelation is independent of the curing rate δ in regular graphs. Moreover, using the Magnus expansion, we are able to calculate the autocorrelation in the transient state of the SIS process. Our analysis of the transient state not only allows the study of the SIS process above or below the epidemic threshold but also possibly applied to the study of the critical behavior [2].

[2]For example, evidence has shown that there exists an extended critical region just above the NIMFA threshold $\tau_c^{(1)}$ in some networks: the Griffiths phase [14, 15, 16] related to the epidemic localization [17, 152, 153, 154] evaluated by the behavior around the NIMFA threshold [155]. With the extended critical region, the effective infection rate τ needs not to be fine-tuned to the exact threshold to let the process be critical [19]. In the critical region, the NIMFA steady-state prevalence $y_\infty^{(1)}\triangleq\lim_{t\to\infty}y^{(1)}(t)$ is very small (converging to zero with the increase of network

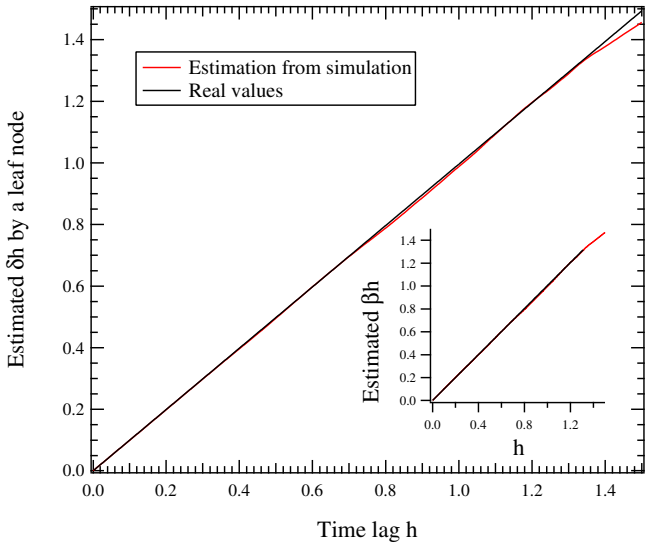

Figure 3.8: The estimation of the infection rate β and the curing rate δ using (3.18) and (3.17) for the star graph corresponding to Fig. 3.3. The curves are $-(1-v_{j\infty})\ln(R_{j\infty}(h))$ and $-\frac{v_{j\infty}\ln(R_{j\infty}(h))}{\sum_{i=1}^{N} a_{ij}v_{i\infty}}$ of a leaf node versus h. Both the estimated β and δ are 1.00, while the real values of rates are both 1.

We also evaluated our results by simulation. Although NIMFA assumes that there is no correlation between the infection states of neighbors, i.e. $E[X_i(t)X_j(t)] = E[X_i(t)]E[X_j(t)]$ for $i \neq j$, we show that the NIMFA autocorrelation ($i = j$) is generally accurate by simulation, and the accuracy depends on the accuracy of the NIMFA infection probabilities. If NIMFA can capture the first-order moments, i.e. the infection probability of each node and the prevalence, under certain SIS parameters and networks, then NIMFA can also be applied to approximate the autocorrelation properties. Finally, we show that our results can be used to estimate the infection and curing rate of the SIS process.

N), and it is similar to that below the threshold $\tau_c^{(1)}$. However, the prevalence is different in the transient state: the prevalence follows a power-law decay in the critical region [15] while it follows an exponential decay below the threshold. The critical autocorrelation properties in the Griffiths phase might be studied with our analysis of the transient state.

4

Non-Markovian SIS Processes

Since a real epidemic process is not necessarily Markovian, the epidemic threshold obtained under the Markovian assumption may be not realistic. To understand general non-Markovian epidemic processes on networks, we study the Weibullian SIS process in which the infection process is a renewal process with a Weibull time distribution. We find that, if the infection rate exceeds $1/\ln(\lambda_1+1)$*, where* λ_1 *is the largest eigenvalue of the network's adjacency matrix, then the infection will persist on the network under the mean-field approximation for any shape of the Weibull infection time distribution. We also show that the same conclusion holds for Gamma and lognormal distributions. Thus,* $1/\ln(\lambda_1+1)$ *is possibly the largest epidemic threshold for a general non-Markovian SIS process with a Poisson curing process under the mean-field approximation. Furthermore, non-Markovian SIS processes may result in a multi-modal prevalence.*[1]

4.1 From Markovian to non-Markovian processes

SINCE real epidemic processes may not be Markovian, non-Markovian epidemic models need to be examined [137]. For simplicity, most research (implicitly) assumes that the networked spreading process is Markovian, which means that both the infection and curing process are Poisson processes. The length of the time interval between two adjacent events (infection or curing in the SIS process) is exponentially distributed in the Poisson process. Under the Markovian assumption, the epidemic threshold τ_c can be approximately obtained by HMF and NIMFA. However, an epidemic process is not necessarily Markovian, and the infection attempts do not happen uniformly with time t as in a Poisson process. For example, the infection time of online information

[1] This chapter is based on [91, 156].

spread follows lognormal distribution [157]. In epidemiology, the infection time T is called generation time [158] which characterizes the infectivity of pathogens and is defined as the time between the infections (or the symptoms onsets) of the primary case and the secondary case infected by the primary case. The generation time is usually obtained by monitoring the first cases and secondary cases in households and follows skewed distributions which can be fitted by the Gamma, Weibull or lognormal distribution. For example, Heijne et al. [159] evaluated a norovirus outbreak [160] in Sweden in 1999 by fitting the generation time with a Gamma distribution. Cowling et al. [161] fitted the generation time of influenza in Hong Kong with all the three distributions and indicated that the Weibull distribution performs slightly better than the Gamma and lognormal distributions based on the Akaike information criterion. Moreover, the generation time of the Severe Acute Respiratory Syndrome (SARS) in Singapore in 2003 is well fitted by the Weibull distribution [162].

Thus, to evaluate the influence of non-Markovian infections, it is reasonable to consider a renewal infection process [88] where the infection time T follows a Weibull, Gamma or lognormal distribution. In the renewal infection process, the distribution of the infection time T, which is the time interval between two adjacent infection attempts of an infected node, is replaced by a more general distribution. In this section, we first discuss the Weibull distribution as in [163, 115],

$$f_T(x) = \frac{\alpha}{b}\left(\frac{x}{b}\right)^{\alpha-1} \mathrm{e}^{-(x/b)^{\alpha}} \tag{4.1}$$

for $x \geq 0$, with the expectation

$$E[T] = b\Gamma\left(1 + \frac{1}{\alpha}\right)$$

where α is a shape parameter, $\Gamma(x)$ is the Gamma function, and $b = (\beta\Gamma(1 + \frac{1}{\alpha}))^{-1}$ because the average infection time $E[T]$ is fixed to the inverse of the infection rate $1/\beta$ in order to compare different α-regimes. Furthermore, the distribution function is

$$F_T(x) = \Pr[T \leq x] = 1 - \mathrm{e}^{-(x/b)^{\alpha}} \tag{4.2}$$

for $x \geq 0$. We refer to this model as a *Weibullian SIS process*. In the Weibullian SIS process, the shape parameter α controls the infection process. The Weibull distribution is heavy-tailed when $\alpha < 1$, exponential when $\alpha = 1$, hence Markovian, and Gaussian-like when $\alpha > 1$ as shown in Fig. 4.1. Furthermore, tuning the shape parameter α dramatically shifts the epidemic threshold, and

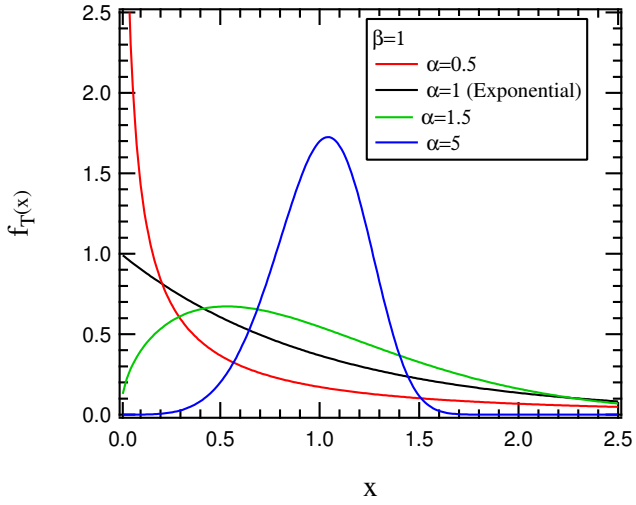

Figure 4.1: The probability density function of Weibull distribution with different value of the shape parameter.

the epidemic threshold increases with the distribution changing from heavy-tailed to Gaussian-like [115, 163].

The Weibullian SIS process is capable of modeling various kinds of non-Markovian epidemic processes by choosing a suitable shape parameter α. For example, the Weibullian SIS process with a heavy-tailed infection time ($\alpha < 1$) predicts a smaller epidemic threshold τ_c, compared to a Markovian SIS process, which agrees with the fact that a heavy-tailed interaction time leads to a longer persistence of infection in reality [164]. However, the shape parameter α is generally not known for a real-life epidemic process, which raises two questions: 1) How small should the effective infection rate τ be to ensure that there is no epidemic on the network? 2) How large should the effective infection rate τ be to ensure a persistence of infection on the network? Obviously, since the Markovian SIS process is a special case ($\alpha = 1$) of the Weibullian SIS process, neither of the answers to those two questions is the NIMFA epidemic threshold $\tau_c^{(1)}$. In this chapter, we make a first step to understand those questions. Under the mean-field approximation, we find that the largest epidemic threshold of the Weibullian SIS process is $\frac{1}{\ln(\lambda_1+1)}$, which is obtained when $\alpha \to \infty$. Since the Weibullian SIS process is able to model a general epidemic process, we argue that the infection can persist on the network when the effective infection rate $\tau > \frac{1}{\ln(\lambda_1+1)}$ for any infection process. Simulation results in Fig. 4.6 seem to support our claim. Moreover, we will show later

that similar results hold for other infection time distributions, i.e. the Gamma and lognormal distributions.

Another motivation of our study is that an infinite shape parameter α leads to a model for synchronized spreading phenomena, which may help us to understand some special cases. For example, a computer virus can be controlled to infect computers periodically, and it is also technically possible for a virus to burst at a same time point. Many computer viruses burst periodically because the developers of a virus spend time on improving the virus before each burst. Thus, the virus development life-cycle and the underlying network collectively determine whether the infection can persist or not. Another example is the seasonal influenza H3N2 where the infection emerges at each influenza season and the prevalence is at a low level between seasons [165]. In those situations, either the infection is synchronized or the infection time interval is sharply Gaussian-like distributed. The Weibullain SIS process with $\alpha \to \infty$ can be applied to approximate those resurgent epidemic processes.

In the following part of this chapter, we first study the Weibullian SIS process with $\alpha \to 0$ to show that the epidemic threshold can be very small. Then, we propose the time-dependent dynamical governing equations for the SIS process with $\alpha \to \infty$ under the mean-field approximation where only the independence between the infection states of neighbors is assumed. We re-evaluate the results by a different non-Markovian mean-field approximation [163] which only applies for the steady-state. Numerical and simulation results are presented for Weibull, Gamma and lognormal distributed infection times to evaluate our theory and to show that all the three different infection processes behave similarly.

4.2 The Weibullian SIS process in the limiting cases

In the Weibullian SIS process, the distribution of the infection time between two adjacent infection attempts of an infected node is a Weibull distribution with an expectation $1/\beta$, and the distribution of the infected time duration is exponential with an expectation $1/\delta$. If $\alpha = 1$, then the Weibullian SIS process reduces to the Markovian SIS process. If the probability of the occurrence of an infection attempt decreases with time, then the process can be modelled by the process with a suitable shape parameter $\alpha < 1$. Otherwise, the infection can be modelled by $\alpha > 1$.

When $\alpha \to 0$, the Weibullian SIS epidemic threshold is 0, and for an arbitrary small α, the mean-field epidemic threshold given in [163] tends to

zero (see Appendix A.3.1).

If $\alpha \to \infty$, then the distribution function (4.2) of the infection time for $T \neq 1/\beta$ tends to

$$\lim_{\alpha\to\infty} F_T(x) = \lim_{\alpha\to\infty} 1 - \mathrm{e}^{-[\beta x\Gamma(1+1/\alpha)]^\alpha} = \begin{cases} 1 & \text{for } x > \frac{1}{\beta} \\ 0 & \text{for } x < \frac{1}{\beta} \end{cases} \tag{4.3}$$

The distribution function is right-continuous, and then $F_T(1/\beta) = \lim_{x\to(1/\beta)^+} F_T(x) = 1$ at the discontinuity $x = \frac{1}{\beta}$. Thus, the probability distribution of the time interval between two adjacent infection events is $\Pr[T = 1/\beta] = 1$ and $\Pr[T \neq 1/\beta] = 0$.

Figure 4.2 shows the time-dependent prevalence $y(t) \triangleq \frac{1}{N}\sum_{i=1}^{N} E[X_i(t)]$, which is the average fraction of the infected nodes in the Weibullian SIS process. Initially, all nodes are infected. For $\alpha \leq 1$, the prevalence $y(t)$ monotonically decreases to the metastable state, and for $\alpha > 1$, the prevalence $y(t)$ fluctuates with a decaying amplitude. When $\alpha \to \infty$, the prevalence $y(t)$ is no longer steady, but periodically changes. There is a huge gap between the maximum and the minimum prevalence. With the increase of α, the amplitude increases, but the minimum prevalence decreases as shown in Fig. 4.2. The persistence of the infection needs a higher effective infection rate τ for a larger α. Figure 4.2 reveals that a non-Markovian infection process may lead to a multi-modal prevalence and infection probability, a function with multiple local maxima over time. The multi-modal prevalence represents the resurgence of the epidemic. Previously, the resurgent phenomenon raised by the underlying structure is discussed in the Susceptible-Infected (SI) [166], the Susceptible-Infected-Recover (SIR) model [167] and the SIS model as in Section 2.5.

As mentioned above, the infected nodes infect their neighbours precisely every $1/\beta$ time unit when $\alpha \to \infty$, and then it is hard to study the process with only one time parameter t as done in the Markovian SIS process. To investigate this process analytically, we divide the time t into time interval of length $1/\beta$ with index $n = 0, 1, \cdots$. The infection state of node j is $X_j(t^* + n/\beta)$ at time t^* of the n-th time interval, where $n \geq 0$ and $t^* \in [0, 1/\beta)$. At $t = 0$, the initially infected nodes are seeded, and the first infection attempt of each infected node happens at the start of the second time interval $t = 1/\beta$. Thus, the infection attempts always happen at the start of each time interval ($t^* = 0$). If a healthy node has an infected neighbour when $t^* \to 1/\beta$, then the healthy node will be infected at the start of the next time interval. The probability that node j is healthy at the end of the n-th time interval and has at least one

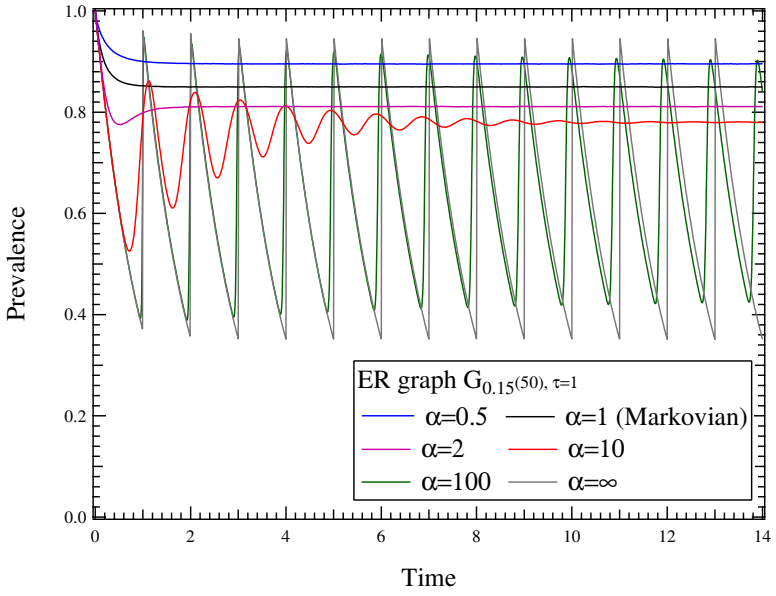

Figure 4.2: The prevalence of the Weibullian SIS process on an Erdős-Rényi (ER) network $G_{0.15}(50)$ obtained by averaging over 10^5 realizations with all nodes infected initially. The effective infection rate $\tau = 1$ is around 8.5 times the NIMFA threshold ($\tau_c^{(1)} = 0.1173$). The minimum prevalence decreases with the shape parameter α. For $\alpha \leq 1$, the prevalence is a trivial unimodal function with time. For small $\alpha > 1$, the prevalence oscillates but eventually becomes approximately constant. When $\alpha \to \infty$, the metastable state prevalence is no longer constant.

infected neighbor is

$$\lim_{t^* \to 1/\beta} \Pr\left[X_j\left(t^* + \frac{n}{\beta}\right) = 0, \sum_{i \in \mathcal{N}_j} X_i\left(t^* + \frac{n}{\beta}\right) \geq 1 \right] \tag{4.4}$$

where $\mathcal{N}_j$ is the set of the neighbours of node j. While the probability that node j is infected at the end of the n-th time interval is

$$\lim_{t^* \to 1/\beta} E\left[X_j\left(t^* + \frac{n}{\beta}\right) \right] \tag{4.5}$$

The probability that node j is infected at the start of the $(n+1)$-th time interval

$E[X_j((n+1)/\beta)]$ is thus the sum of (4.4) and (4.5), and we obtain,

$$\begin{aligned} & E\left[X_j\left(\frac{n+1}{\beta}\right)\right] \\ & = \lim_{t^*\to 1/\beta}\left(\Pr\left[X_j\left(t^*+\frac{n}{\beta}\right)=0, \sum_{i\in\mathcal{N}_j} X_i\left(t^*+\frac{n}{\beta}\right)\geq 1\right]\right. \\ & \left. + E\left[X_j\left(t^*+\frac{n}{\beta}\right)\right]\right) \end{aligned} \tag{4.6}$$

Equation (4.6) is not analytically solvable. Here, we apply a mean-field approximation to solve (4.6), similar as in NIMFA for the Markovian SIS process. We assume that the infection state between neighbours is independent at the end of each time interval, i.e., $\lim_{t^*\to 1/\beta} E[X_i(t^*+n/\beta)X_j(t^*+n/\beta)] = \lim_{t^*\to 1/\beta} E[X_i(t^*+n/\beta)]E[X_j(t^*+n/\beta)]$. Under this assumption, we denote the approximate value of $E[X_j(t)]$ by $v_j(t)$, and the infection probabilities by a column vector $\mathbf{v}(t) \triangleq [v_1(t), \cdots, v_N(t)]^T$. Thus, the mean-field infection probability at $t^*=0$ of the $(n+1)$-th time interval follows,

$$\begin{aligned} v_j\left(\frac{n+1}{\beta}\right) = & \lim_{t^*\to 1/\beta}\left(\left[1-v_j\left(t^*+\frac{n}{\beta}\right)\right]\left(1-\prod_{i\in\mathcal{N}_j}\left[1-v_i\left(t^*+\frac{n}{\beta}\right)\right]\right)\right. \\ & \left. + v_j\left(t^*+\frac{n}{\beta}\right)\right) \end{aligned} \tag{4.7}$$

In each time interval, an infected node can be cured at any time point with a equal probability during $t^* \in [0, 1/\beta)$, because the curing process is Poissonian. The governing equation of the infection probability of node j for $j = 1, \cdots, N$ is,

$$\frac{\mathrm{d}v_j\left(t^*+n/\beta\right)}{\mathrm{d}t^*} = -\delta v_j\left(t^*+\frac{n}{\beta}\right)$$

for $t^* \in [0, 1/\beta)$. Given the initial condition $v_j(n/\beta)$, the solution of the equation above is

$$v_j\left(t^*+\frac{n}{\beta}\right) = v_j\left(\frac{n}{\beta}\right)\mathrm{e}^{-\delta t^*} \tag{4.8}$$

for $t^* \in [0, 1/\beta)$.

Substituting (4.8) evaluated at $t^* \to 1/\beta$ of the n-th time interval, thus $\lim_{t^* \to 1/\beta} v_j\left(t^* + \frac{n}{\beta}\right) = v_j(n/\beta)\mathrm{e}^{-1/\tau}$ into (4.7), we obtain a recursion of the infection probability at $t^* = 0$ of each time interval,

$$v_j\left(\frac{n+1}{\beta}\right) = \left[1 - v_j\left(\frac{n}{\beta}\right)\mathrm{e}^{-1/\tau}\right]\left(1 - \prod_{i \in \mathcal{N}_j}\left[1 - v_i\left(\frac{n}{\beta}\right)\mathrm{e}^{-1/\tau}\right]\right) + v_j\left(\frac{n}{\beta}\right)\mathrm{e}^{-1/\tau} \tag{4.9}$$

Equation (4.9) has a similar form as the discrete-time SIS process, which has been studied in [168]. In the metastable state $n \to \infty$, the infection probability $v_j(n/\beta)$ at the start of each time interval $t^* = 0$ is constant. We can check whether the infection probability $\lim_{n\to\infty} v_j(n/\beta)$ is zero or not, to obtain the epidemic threshold. Consequently, we arrive at the following result.

Theorem 4.2.1 The mean-field epidemic threshold of the Weibullian SIS process on a connected network with $\alpha \to \infty$ obtained by (4.8) and (4.9) is

$$\tau_c^{(B)} = \frac{1}{\ln(1 + \lambda_1)} \tag{4.10}$$

If the effective infection rate $\tau > \tau_c^{(B)}$, then infection can persist on the network with a non-zero steady periodic infection probability $\mathbf{v}_\infty(t^*) \triangleq \lim_{n\to\infty} \mathbf{v}(t^* + n/\beta)$, and $\mathbf{v}_\infty(t^*) = \mathbf{v}_\infty(0)\mathrm{e}^{-\delta t^*}$ for $t^* \in (0, 1/\beta]$. If $\tau < \tau_c^{(B)}$, then the epidemic process enters the all-healthy state in the long run $\lim_{t\to\infty} \mathbf{v}(t) = \mathbf{0}$.

The proof of Theorem 4.2.1 is in Appendix A.3.2. The superscript (B) in $\tau_c^{(B)}$ refers to *Burst*. The epidemic threshold (4.10) has a similar form as the NIMFA epidemic threshold [102] and the discrete-time SIS [129, 128] threshold $1/\lambda_1$, but with a logarithmic relation to the largest eigenvalue λ_1 of the adjacency matrix A of the network. Furthermore, the term $1 + \lambda_1$ in the logarithmic function ensures that the epidemic threshold (4.10) is positive for $\lambda_1 > 0$ in any finite-size connected network. The threshold (4.10) of a scale-free network with a finite average degree [169, 92] converges to zero in the thermodynamic limit $N \to \infty$.

When the effective infection rate $\tau < \tau_c^{(B)}$, the infection probability $v_j(t)$ of each node decreases to zero in the long run. We represent $x_i \leq y_i$ and $x_i < y_i$ for all i by the vector relationship $[x_1, \cdots, x_n]^T \preceq [y_1, \cdots, y_n]^T$ and $[x_1, \cdots, x_n]^T \prec [y_1, \cdots, y_n]^T$, respectively. If $\tau < \tau_c^{(B)}$, then the infection probability $\mathbf{v}(t)$ is upper bounded by an exponentially decreasing function with time t, which is

$$\mathbf{v}(t) \prec \left(\mathrm{e}^{-\delta}(\lambda_1 + 1)^{\beta}\right)^t \mathbf{z}$$

where $\mathbf{z}$ is a constant vector whose every element is positive (see Appendix A.3.3). Furthermore, the mean-field prevalence $y^{(1)}(t) \triangleq \frac{1}{N}\sum_{i=1}^{N} v_i(t)$ is upper bounded by $y^{(1)}(t) < \left(\mathrm{e}^{-\delta}(\lambda_1 + 1)^{\beta}\right)^t c$, where c is a positive value.

When $\tau > \tau_c^{(B)}$, the steady infection probability reaches a maximum $\mathbf{v}_\infty(0)$ at the start of each time interval $t^* = 0$, and a minimum $\mathbf{v}_\infty(0^-) \triangleq \lim_{t^* \to 1/\beta} \mathbf{v}_\infty(t^*)$ at the end of each time interval $t^* \to 1/\beta$. The steady maximum infection probability $\mathbf{v}_\infty(0)$ can be obtained by solving (4.9) numerically. Since $\mathbf{v}_\infty(0) = \mathbf{v}_\infty(0^-)\mathrm{e}^{1/\tau}$, the ratio between the maximum and minimum steady infection probability is,

$$\frac{\mathbf{v}_\infty(0)}{\mathbf{v}_\infty(0^-)} = \mathrm{e}^{1/\tau} < \lambda_1 + 1 \tag{4.11}$$

The last inequality holds because the effective infection rate τ is above the mean-field threshold $\tau > \tau_c^{(B)}$. The inequality (4.11) indicates that the burst of the infection in the steady state is restricted by the underlying network, specifically, the largest eigenvalue λ_1 of the adjacency matrix A.

4.3 Numerical and simulation results

4.3.1 The limiting case

We evaluate the mean-field method by comparing the approximation with the simulation of the exact Weibullian SIS process. The simulation are performed on an ER network, a scale-free network, and a rectangular grid network.

Figure 4.3, 4.4 and 4.5 present the prevalence of the Weibullian SIS process with $\alpha \to \infty$ in the long run, together with the NIMFA and the Markovian prevalence. Steady maximum prevalence and minimum prevalence under the

mean-field approximation and simulation are shown. The prevalence is obtained by averaging over 10^5 realizations of simulation with all nodes infected initially to prevent the inaccuracy caused by the early die-out [26]. The simulation runs for a long enough time (50 time units with $\delta = 1$), and the maximum and minimum prevalence are plotted, which are selected from the last complete time period.

The numerical solution of Eq. (4.8) and (4.9) approximates the simulation results well, and the phase transition of the simulated process happens around the mean-field threshold $\tau_c^{(B)}$. Among all the three different networks, the accuracy of the mean-field approximation is worst in the rectangular grid network with a minimum largest eigenvalue $\lambda_1 = 3.9627$, and best in the scale free-network with a largest $\lambda_1 = 11.9944$. The simulations also show that the the mean-field threshold $\tau_c^{(B)} = \frac{1}{\ln(1+\lambda_1)}$ is, just as for NIMFA, a lower bound.

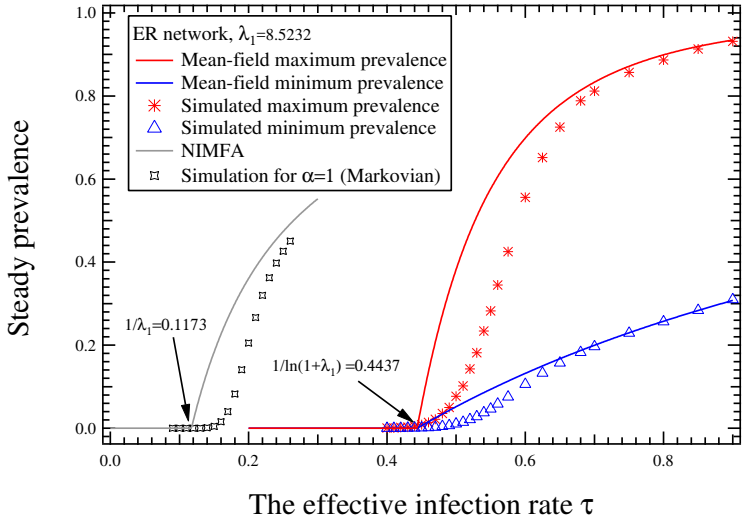

Figure 4.3: The ER network $G_{0.15}(50)$ corresponding to Fig. 4.2.

4.3.2 The range of the epidemic threshold

The marks in Fig. 4.6 are the epidemic threshold of the Weibullian SIS process with different shape parameter α. The simulation result is partly reproduced in [170] by an independent simulation, based on [133]. As mentioned above, the epidemic threshold can be approximate zero, which agrees with the simulation results. With the increase of α, the epidemic threshold τ_c converges

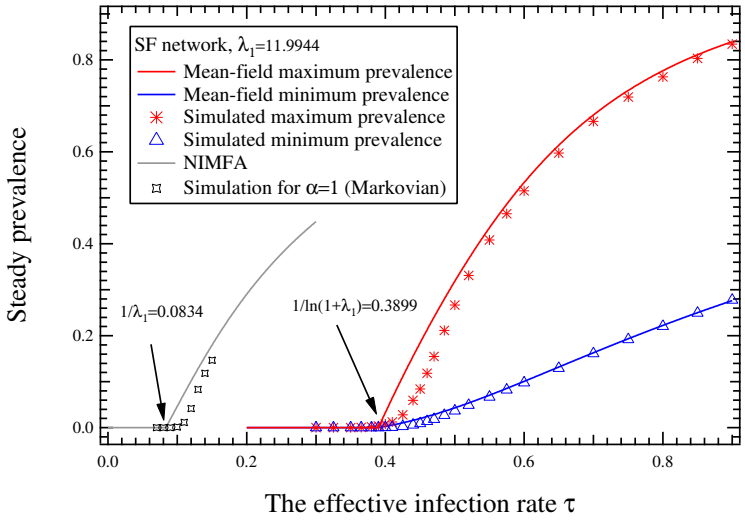

Figure 4.4: A Barabási-Albert scale-free network with size $N = 500$, and number of links $L = 1491$.

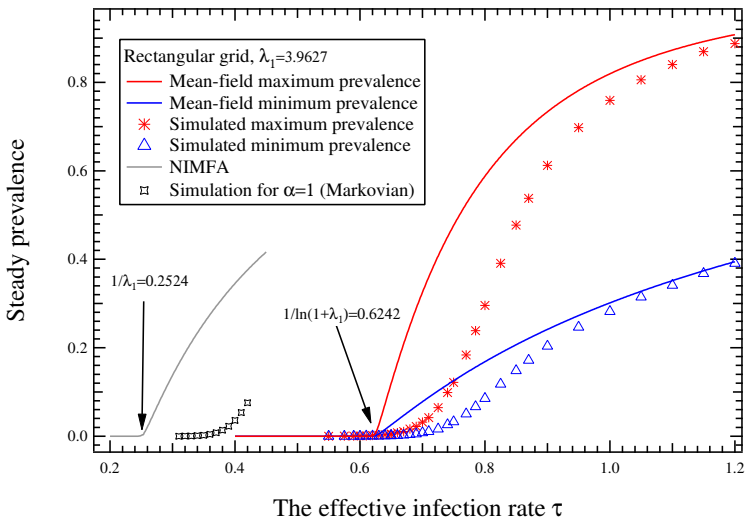

Figure 4.5: A rectangular grid with size $N = 484$, $L = 924$.

approximately to $\tau_c^{(B)}$.

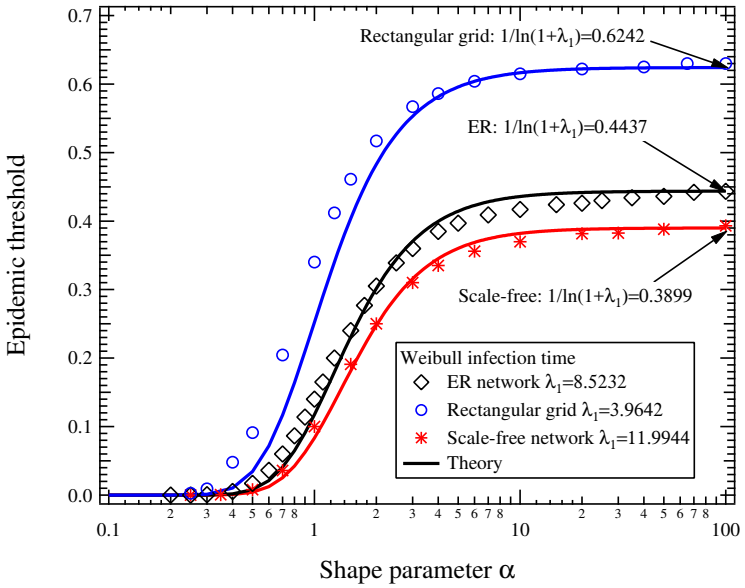

Figure 4.6: The epidemic threshold versus the Weibull shape parameter α. The thresholds are obtained by simulation of 10^5 realizations. The simulation setup is same as that in Fig. 4.3, 4.4 and 4.5. The threshold is chosen as the value of the effective infection rate τ which leads to the maximum prevalence being around 0.001 at the last period.

4.4 Rethinking the threshold with renewal theory

For other distributions of infection time, the SIS process behaves similarly. In this section, we discuss the epidemic threshold further by a non-Markovian mean-field approximation base on the renewal theory.

4.4.1 Non-Markovian mean-field approximation based on renewal theory

Assuming that the constant metastable state exists and invoking renewal theory, the mean-field steady state probability $v_{i\infty}$ of infection of node i is shown

in [163] to obey

$$E[M](1 - v_{i\infty}) \sum_{j=1}^{N} a_{ij} v_{j\infty} = v_{i\infty}, \tag{4.12}$$

which is exactly the same equation as in NIMFA in the exponential case, if we replace the effective infection rate $\tau = \beta/\delta$ by the average number $E[M]$ of infection events during a healthy period, which is [163]

$$E[M] = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\varphi_T(z)\,\varphi_R(-z)}{1 - \varphi_T(z)} \frac{dz}{z} \tag{4.13}$$

where $\varphi_T(z) = E\left[e^{-zT}\right]$ and $\varphi_R(z)$ are the probability generating function (pgf) of the infection time T and recovery or curing time R, respectively.

The analogy with the NIMFA equations (2.4) immediately leads to the epidemic threshold in non-Markovian SIS epidemics

$$E[M_c] = \frac{1}{\lambda_1} \tag{4.14}$$

where λ_1 is the largest eigenvalue of the adjacency matrix A of the graph G. Thus, if $E[M] > \frac{1}{\lambda_1}$, then the epidemic process is eventually endemic (in the mean-field approximation), in which a non-zero fraction of the nodes remain infected. If $E[M] < \frac{1}{\lambda_1}$, then the epidemic process dies out and the network is overall healthy in the long run. As shown in [163], the mean-field epidemic threshold is the effective infection rate $\tau = \beta/\delta$ using the solution of

$$\phi_T(\delta) = \frac{1}{\lambda_1 + 1} \tag{4.15}$$

when the curing is a Poisson process.

Here, the renewal theory assumes a constant steady-state prevalence $v_{i\infty}$ in (4.12) which does not exist when $\alpha \to \infty$ as the previous analysis in this chapter, where the initially infected nodes are seeded synchronously, indicated. However, the renewal theory can still lead to the epidemic threshold $\tau_c^{(B)} = \frac{1}{\ln(1+\lambda_1)}$ and $v_{i\infty}$ can be interpreted as the infection probability over all possible initial conditions where the seeding of initially infected can be asynchronous.

4.4.2 The Weibull infection time

Apart from the time-dependent analysis of the limiting case $\alpha \to \infty$ of the Weibullian SIS process, we now reconsider the non-Markovian SIS process with renewal theory provided in the previous section 4.4.1.

First, the limit case $\alpha \to \infty$ is immediate from (A.23) and (4.15)

$$\frac{1}{1+\lambda_1} = \lim_{\alpha\to\infty} \int_0^\infty e^{-\frac{u^{\frac{1}{\alpha}}}{\Gamma\left(1+\frac{1}{\alpha}\right)\tau}-u} du = \int_0^\infty e^{-\frac{1}{\tau}-u} du = e^{-\frac{1}{\tau}}$$

so that the solution is $\ln(1+\lambda_1) = \tau^{-1}$. Thus, we establish the result

$$\lim_{\alpha\to\infty} \tau_c^{(B)}(\alpha) = \frac{1}{\ln(1+\lambda_1)} \tag{4.16}$$

derived differently.

A second derivation interprets the general equation (4.14) directly, without resorting to the integral representation in (4.13). For $\alpha \to \infty$, the average number $E[M]$ of infection events during a healthy period R is computed as follows. Without loss of generality, we assume that a node i is infected at time $t = 0$. The infected node i infects its neighbors at times $t = 1/\beta, \ldots, k/\beta, \ldots$, until node i is recovered at time R. The recovery time R follows an exponential distribution with expectation $1/\delta$. Thus, if the recovery time R falls into $k/\beta < R < (k+1)/\beta$, then the infected node i infects each of its neighbor k times and we obtain

$$E[M] = \sum_{k=0}^{\infty} k \int_{k/\beta}^{(k+1)/\beta} \delta e^{-\delta x} dx = \sum_{k=0}^{\infty} k \left(e^{-k/\tau} - e^{-(k+1)/\tau}\right) = \frac{1}{e^{\frac{1}{\tau}} - 1}$$

Equating $\frac{1}{e^{\frac{1}{\tau}}-1} = 1/\lambda_1$ in (4.14) again leads to (4.16).

For general α, the pgf of a Weibull random variable T is given in (A.22) in Appendix A.3.4,

$$\varphi_T(w;\alpha) = \alpha \int_0^\infty e^{-wx-x^\alpha} x^{\alpha-1} dx$$

where $w = bz$. Using criterion (4.15), the mean-field epidemic threshold (under the renewal theory) $\tau_{c;W}^{(1)}(\alpha)$ in (4.14) of the non-Markovian SIS process with Weibullian infection time T is the solution for τ in

$$\varphi_T\left(\frac{1}{\Gamma\left(1+\frac{1}{\alpha}\right)\tau}; \alpha\right) = \frac{1}{1+\lambda_1} \tag{4.17}$$

The direct numerical solution of (4.17) fits the epidemic threshold very well as shown in Fig. 4.6. Explicit solution of (4.17) based on Langrange series and asymptotic expansion can be found in [91].

4.4.3 The Gamma infection time and its interpretation

Instead of a Weibull distribution, we also consider a Gamma distribution [88, pp. 45-46] for the infection time T,

$$f_{T_{\text{Gamma}}}(x;\xi) = \frac{\frac{1}{b_\Gamma}\left(\frac{x}{b_\Gamma}\right)^{\xi-1}}{\Gamma(\xi)}\mathrm{e}^{-\frac{x}{b_\Gamma}} \tag{4.18}$$

with $E[T_{\text{Gamma}}] = b_\Gamma\xi$ and the corresponding pgf is

$$\varphi_{T_{\text{Gamma}}}(z;\xi) = (1+b_\Gamma z)^{-\xi}. \tag{4.19}$$

Similar to the Weibull distribution, the Gamma distribution reduces for $\xi = 1$ to an exponential distribution and here ξ is the shape parameter as α in the Weibull distribution. After fixing the average infection $E[T]$ to $\frac{1}{\beta}$, the value of $b_\Gamma = \frac{1}{\beta\xi}$. Using criterion (4.15), the mean-field epidemic threshold for general ξ is

$$\tau_{c;\Gamma}^{(1)}(\xi) = \frac{1}{\xi\left((1+\lambda_1)^{\frac{1}{\xi}}-1\right)}. \tag{4.20}$$

For the limiting case $\xi \to \infty$, after substitution $x = \frac{1}{\xi}$ we have

$$\lim_{x\to 0}\tau_{c;\Gamma}^{(1)}(\xi^{-1}) = \lim_{x\to 0}\frac{x}{((1+x)^x-1)} = \frac{1}{\ln(1+\lambda_1)} \tag{4.21}$$

Figure 4.7 shows the epidemic threshold vs the shape parameter ξ. The theoretical results well fit the simulation.

An interpretation of the Gamma infection time

If $\xi = k \geq 1$ is an integer, then the Gamma random variable equals the sum of k independent and identically distributed exponential random variable [88, pp. 45-46]. Thus, the SIS model with a Gamma infection time T can be interpreted as a *dose-infection* process: Each infected node can infect each healthy neighbor via a Poisson process with rate r but only a small dose of infection is transmitted. A healthy node needs to receive $k \geq 1$ continuous doses of infection from an infected neighbor to become infected. Under this interpretation, the infection time T follows a Gamma distribution with $\xi = k$ and $b_\Gamma = 1/r$. The overall effective infection rate $\tau = 1/(E[T]\delta) = r/(k\delta)$. Thus, there exists a *dose threshold* k_c such that if $k > k_c$, then the SIS process

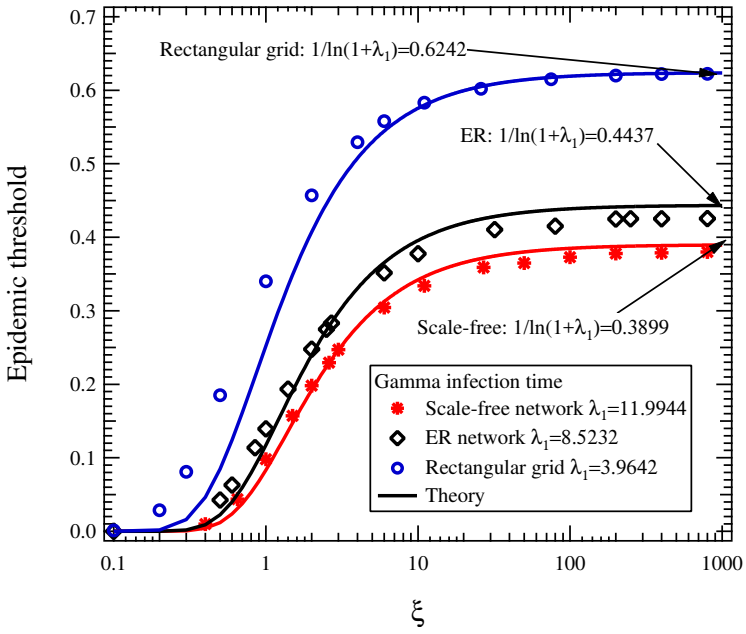

Figure 4.7: The epidemic threshold of the SIS model with Gamma infection time for the three different networks. The full line is (4.20) while the marks are obtained from simulation.

is below the epidemic threshold while if $k < k_c$, then above the threshold. Here, the dose threshold k_c can be a real number. Equating $\tau^{(1)}_{c;\Gamma}(\xi) = r/(k_c\delta)$ in (4.20) with $\xi = k_c$, we obtain the following dose threshold

$$k_c = \frac{\log(1+\lambda_1)}{\log(\delta/r+1)}. \tag{4.22}$$

Equation (4.22) shows that the dose threshold k_c increases logarithmically with the largest eigenvalue λ_1 of the underlying network. When $r > 1$, then $\log(1+\frac{\delta}{r}) \approx \frac{\delta}{r} + O(\frac{1}{r^2})$ which lead to $k_c \approx \log(1+\lambda_1)\frac{r}{\delta}$ for sufficiently large r: the dose threshold k_c increases approximately linearly with the transmission rate r of each dose of infection.

4.4.4 The lognormal infection time

The lognormal infection time follows

$$f_{T_{\text{lognormal}}}(x) = \frac{\exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right)}{\sigma x\sqrt{2\pi}} \tag{4.23}$$

with mean $E[T_{\text{lognormal}}] = \exp(\mu + \frac{\sigma^2}{2})$ and pgf

$$\varphi_{T_{\text{lognormal}}}(z;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty \mathrm{e}^{-zx}\frac{\mathrm{e}^{-\frac{(\log x-\mu)^2}{2\sigma^2}}}{x}dx \tag{4.24}$$

Let $E[T_{\text{lognormal}}] = \frac{1}{\beta}$ and using the threshold criterion (4.15), we have the equation of τ

$$\frac{1}{\lambda_1+1} = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty \mathrm{e}^{-\frac{y}{\tau}}\frac{\mathrm{e}^{-\frac{(\log y+\frac{\sigma^2}{2})^2}{2\sigma^2}}}{y}dy \tag{4.25}$$

whose solution is the epidemic threshold $\tau_{c;l}^{(1)}(\sigma)$.

Since $\lim_{\sigma\to 0} f_{T_{\text{lognormal}}}(x) = \delta_D(x-e^{\mu}) = \delta_D(x-\frac{1}{\beta})$ where $\delta_D(x)$ is the Dirac delta function, the threshold criterion (4.15) with lognormal infection time T becomes

$$\frac{1}{\lambda_1+1} = \int_0^\infty \mathrm{e}^{-\delta y}\delta_D\left(y-\frac{1}{\beta}\right)dy = \mathrm{e}^{-\frac{1}{\tau}} \tag{4.26}$$

Thus, it holds that $\lim_{\sigma\to 0}\tau_{c;l}^{(1)}(\sigma) = \lim_{\alpha\to\infty}\tau_{c;W}^{(1)}(\alpha) = \lim_{\xi\to\infty}\tau_{c;\Gamma}^{(1)}(\xi)$ which further supports the claim that $\tau_c^{(B)} = \frac{1}{\ln(\lambda_1+1)}$ is a maximum possible SIS epidemic threshold. The simulation shown in Fig. 4.8 exhibit a similar behavior for the lognormal epidemic threshold $\tau_{c;l}^{(1)}(\sigma^{-1})$ as for $\tau_{c;W}(\alpha)$ and $\tau_{c;\Gamma}^{(1)}(\alpha)$

4.5 Conclusion

As a general model, the Weibullian SIS process can model a general non-Markovian SIS process by choosing a suitable shape parameter α. We study the process in the extreme situation $\alpha \to 0$ and ∞ to obtain an understanding of the influence of the underlying network on a general epidemic process. For

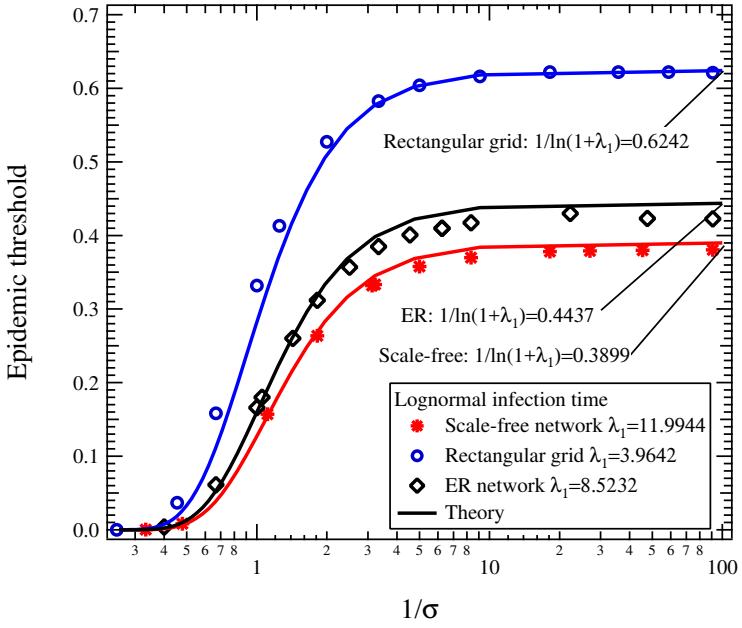

Figure 4.8: The epidemic threshold of the SIS model with lognormal infection time for the three different networks. The full line is the numerical solution of (4.25) while the marks are obtained from simulations.

an SIS process with an unknown infection process, our results reveal that the certainty about the extinction of infection is not possible even if the effective infection rate τ is small, but the infection can always persist on the network if the effective infection rate $\tau > 1/\ln(\lambda_1 + 1)$. Additionally, we obtain the properties of the synchronized epidemic process, i.e., the Weibullian SIS process with $\alpha \to \infty$ (also for the lognormal infection time $\sigma \to 0$ and for the Gamma infection time $\xi \to \infty$). Our results derived by the time-dependent mean-field analysis are in line with those from the previous mean-field theory based on renewal theory. For other realistic distributions of infection time, i.e. the Gamma and lognormal distributions, the SIS processes are shown to behave similarly.

5

Network Localization of Spreading Processes

To shed light on the disease localization phenomenon, we study a bursty SIS model[1] *and analyze the model under the mean-field approximation. In the bursty SIS model, the infected nodes infect all their neighbors periodically, and the near-threshold steady-state prevalence is non-constant and maximized by a factor equal to the largest eigenvalue λ_1 of the adjacency matrix of the network. We show that the maximum near-threshold prevalence of the bursty SIS process on a localized network tends to zero even if λ_1 diverges in the thermodynamic limit, which indicates that the burst of infection cannot turn a localized spreading into a delocalized spreading. Our result is evaluated both on synthetic and real networks.*[2]

5.1 The SIS localization: a near-threshold behavior

THE near-threshold behavior, i.e. the behavior around the threshold where a phase transition occurs, is of great interest in the study of dynamical processes, because many real complex systems may operate near the phase transition point [172, 173, 174, 175]. For some networks, the SIS epidemic remains restricted into a small subnetwork and does not spread over the whole network for infection strength just above the (mean-field) epidemic threshold. This restricted spreading phenomenon is known as the (metastable) localiza-

[1] The limiting case of the Weibullian SIS model in Chapter 4 where the distribution of the infection time is a Dirac delta function.

[2] This chapter is based on [171].

tion[3] of the SIS process [17, 154, 181], and has been studied recently. de Arruda *et al.* [181] investigated the localization phenomenon of SIS processes on multiplex networks. Sahneh *et al.* [153] focused on the localization by a maximum entropy and optimization approach. Another near-threshold behavior, called *Griffiths phase*[4] of the SIS process, which is related to localization, is studied by Cota *et al.* [15] and Muñoz *et al.* [16]. The near-threshold behavior of the SIS process has also been applied to explain the operation of brain [19].

In this chapter, we further study the SIS localization phenomenon. In previous studies [17, 154], localization of epidemic processes means that only a finite number of nodes is infected in the thermodynamic limit, i.e. when the network size $N \to \infty$. In this work, the definition of epidemic localization is that the average fraction of infected nodes, i.e. the prevalence, tends to zero in the thermodynamic limit, but the number of infected nodes is not necessarily finite. In the following part, we first clarify some misconceptions about the SIS localization in previous studies and show the availability of mean-field methods [17, 154, 102, 108]. We point out that the order of the near-threshold prevalence as a function of the network size N is essential for understanding the influence of the network structure on spreading processes. Motivated by the essence of the prevalence order, we confine ourselves to a mean-field approximation and study a bursty spreading effect which maximizes the near-threshold prevalence by a factor equal to the largest eigenvalue λ_1 of the adjacency matrix of the network. Even though the spectral radius λ_1 diverges with network size N, the spreading bursts cannot change a localized spreading to a delocalized one if the principal eigenvector of the adjacency matrix of the network is localized.

5.2 SIS and eigenvector localizations

In the SIS process, the whole network can be in two different phases in the steady or metastable state: (a) in the all-healthy phase or (b) in the endemic

[3] Localization also appears in other physical systems with different characterizations [176, 98, 177, 178, 179, 180].

[4] The terminology *Griffiths phase* is borrowed from the study of Ising ferromagnet. Griffiths finds that the magnetization of a random Ising ferromagnet is not an analytic function of external field H at $H = 0$ between the critical temperatures of the random and the corresponding pure Ising ferromagnet [182], but in the study of epidemic processes, the non-analyticity of the function of the prevalence just above the epidemic threshold in the thermodynamic limit is still unknown.

phase. In the all-healthy phase, the epidemic has disappeared. In the endemic phase, the infection can persist in the network. The SIS process experiences the phase transition at the threshold τ_c, which can be determined by the mean-field method $\tau_c^{(1)} = 1/\lambda_1$. If the effective infection rate $\tau > \tau_c^{(1)}$, then the process is in the endemic phase under mean-field theory; otherwise, in the all-healthy phase.

For a finite network, the endemic and all-healthy phases can be identified by the prevalence, which can be considered as an order parameter for the SIS process. A non-zero prevalence implies the endemic phase and a zero prevalence means the all-healthy phase. However, in the thermodynamic limit where the network size $N \to \infty$, a zero prevalence does not necessarily coincide with an all-healthy state just above the epidemic threshold. Goltsev *et al.* [17] considered the zero prevalence in the thermodynamic limit as an indication of the localization phenomenon of the SIS process, where only a finite number of nodes are infected on average. In particular, Goltsev *et al.* [17] evaluate the steady-state prevalence $y_\infty(\tilde{\tau}) = \frac{1}{N}\sum_{i=1}^{N} v_{i\infty}(\tilde{\tau})$ just above the mean-field epidemic threshold by its first-order expansion $y_\infty(\tilde{\tau}) = a\tilde{\tau} + o(\tilde{\tau})$ with [155]

$$a = \frac{\sum_{i=1}^{N} x_i}{N\sum_{i=1}^{N} x_i^3} \tag{5.1}$$

where x_i is the ith component of the principal eigenvector of the adjacency matrix, obeying the normalized condition $\sum_{i=1}^{N} x_i^2 = 1$ and $\tilde{\tau} \triangleq \tau/\tau_c^{(1)} - 1 \ll 1$ is the normalized effective infection rate. A tight bound of a is $\frac{\min_i x_i}{\max_i x_i} < a < \frac{1}{\min_i x_i \sqrt{N}}$ as derived in Appendix B. If $a \to 0$ as $N \to \infty$, then the near-threshold prevalence is zero, and if $a > 0$ as $N \to \infty$, then a non-zero fraction of nodes are infected just above the threshold. Goltsev *et al.* [17] define localization by the inverse participant ratio (IPR) $\eta(x) = \sum_{i=1}^{N} x_i^4$ of the principal eigenvector x, and state that if the IPR $\eta(x) = O(1)$, then the principal eigenvector x is localized in a few components $x_i = O(1)$ and only a finite number of nodes are infected in the network with $a \to 0$ as $N \to \infty$. Otherwise, if $\eta(x) = o(1)$, then the vector x is delocalized such that each component $x_i = O(\frac{1}{\sqrt{N}})$. Ferreira *et al.* [154] argue that if a finite number of nodes are infected using mean-field theory, then the virus eventually dies out and then the mean-field approximations [12, 102] fail due to their omission of the absorbing state.

However, a zero prevalence in the thermodynamic limit does not necessarily mean that the number of infected nodes is finite. To illustrate this fact, let us consider a scale-free network which follows a power-law degree dis-

tribution with exponent γ, i.e. $\Pr[D = k] = \frac{k^{-\gamma}}{\zeta(\gamma)}$, $k \in \mathbb{N}$ and $\zeta(\gamma)$ is the Riemann zeta function [183], in the thermodynamic limit. If the average degree of a scale-free network is finite, then $\gamma > 2$ for $N \to \infty$, because $E[D^m] = \zeta(\gamma - m)/\zeta(\gamma)$ converges when $\gamma > m + 1$. The maximum degree scales as $d_{\max} = O(N^{1/(\gamma-1)})$ as derived in [88, p. 594], and thus we may find nodes with degree $O(N^\alpha)$ for $\alpha < 1/(\gamma - 1)$. Given a constant c, the expected number of nodes $\bar{n}_d$ with degree $d = [cN^\alpha]$ is $\bar{n}_d = N \Pr[D = [cN^\alpha]] = (c^{-\gamma} N^{1-\alpha\gamma})/\zeta(\gamma)$. If $0 < \alpha < \frac{1}{\gamma}$, then $\lim_{N\to\infty} \bar{n}_d = \infty$. Thus, the average number of hubs diverges. For each hub with degree of the order $O(N^\alpha)$ for $\alpha > 0$, the local star subgraph ensures that the infection can persist for the effective infection rate $\tau > 0$ in the thermodynamic limit. Related discussions can be found in [184, 185], where the epidemic threshold of power-law networks is shown to be zero in the thermodynamic limit.

Furthermore, the principal eigenvector x may not be localized in a finite subgraph, but localized in a subgraph whose size increases as $O(N^\alpha)$ with $0 < \alpha < 1$ with N. Pastor-Satorras and Castellano [186, 152] define the vector x to be delocalized, only when the IPR $\eta(x) = O(N^{-1})$, while if $\eta(x) = O(N^{-\alpha})$ with $0 \leq \alpha < 1$, then x is localized on a subgraph of size order of $O(N^\alpha)$. An example that can be exactly evaluated is the star-like, two-hierarchical graph [90, p. 143]. In this graph, there are m fully connected nodes, and each node as hub is connected to m leaf nodes. Basically, the graph consists of m fully meshed m-stars. The network size is $N = m^2 + m$ and the average degree is $d_{av} = 3 - \frac{4}{m+1} \approx 3$ for a large network. The largest eigenvalue λ_1 of the graph is m as derived in [90, p. 145], which is actually well approximated by the degree of each node in the maximum K-core [187]. One may verify that the principal eigenvector

$$x = \left[\underbrace{\frac{1}{\sqrt{m+1}}, \ldots, \frac{1}{\sqrt{m+1}}}_{m}, \underbrace{\frac{1}{m\sqrt{m+1}}, \ldots, \frac{1}{m\sqrt{m+1}}}_{m^2} \right]^T$$

is localized on a clique with size in the order of $O(1/\sqrt{N})$ and the IPR $\eta(x) = O(N^{-0.5})$. In this graph, the coefficient $a = O(\frac{1}{\sqrt{N}})$ leads to a zero prevalence, but the average number of infected nodes $N y_\infty(\tilde{\tau}) = O(\sqrt{N})$ diverges in the thermodynamic limit.

Even if the principal eigenvalue x is localized in a finite subgraph and the IPR $\eta(x) = O(1)$, the average number of infected nodes may not be finite in

the thermodynamic limit. Let us consider the extreme case of a star graph, whose principal eigenvector is $x = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2(N-1)}}, \ldots, \frac{1}{\sqrt{2(N-1)}}]^T$. We may verify that the IPR $\eta(x) = O(1)$ and the coefficient $a = O(1/\sqrt{N})$. The average number of infected nodes is $Ny_\infty(\tilde{\tau}) = O(\sqrt{N})$. Thus, just above the epidemic threshold (see also [140] for an exact, asymptotic analysis), an infinite number of nodes is infected, but the prevalence $y_\infty(\tilde{\tau}) = O(\frac{1}{\sqrt{N}})$ tends to zero in the thermodynamic limit.

Our conclusions are: a) the localization of the principal eigenvector and the SIS epidemic process are related, but do not exactly correspond, because the infection can persist in subgraphs which correspond to the delocalized parts of the principal eigenvector; b) a zero prevalence just above threshold in the thermodynamic limit does not imply that the number of infected nodes is finite. Even for the star graph, the average number of infected nodes is of order $O(\sqrt{N})$ just above the epidemic threshold. Thus, it might be impossible to find a network, where the near-threshold number of infected nodes is finite in the thermodynamic limit under the mean-field theory. We address those conclusions to show that: a) in the thermodynamic limit, mean-field theories are consistent and applicable to study the near-threshold behavior because the epidemic may never die out; b) the order of the prevalence as a function of the network size N is essential in the near-threshold spreading dynamic, which is also the motivation of our work. In the following part, we consider a network localized if the IPR $\eta(x) = O(N^{-\alpha})$ for $0 \leq \alpha < 1$, and is delocalized only if $\eta(x) = O(N^{-1})$ as defined by Pastor-Satorras and Castellano [186, 152].

Throughout this chapter, we confine ourselves to the mean-field method. Beyond the mean-field theory, the correlation between infection states of neighbors needs to be taken into consideration. In some cases, the correlation can be substantial. For example, the covariance of the infection state between neighbors in an infinite cycle graph is shown [142, Theorem 3] to be $\xi = 0.121375$ which is apparently not negligible and may introduce long-range correlations. The effect of long-range correlations on localization is unclear and the understanding of localization beyond mean-field theories is still open.

5.3 Evaluating localization under bursts of infection

Since our focus lies on the order of the prevalence as a function of network size N, we construct an SIS process with a non-constant prevalence in the

steady state. We consider bursts that infect all healthy neighbors, leading to an explosion of the spreading. We choose periodical infections to allow analysis, and confine the SIS process to an infectious regime just above the epidemic threshold by tuning the period of the bursts. In some heterogeneous networks, e.g. scale-free networks, the ratio between the maximum prevalence (after each burst) and the minimum prevalence (before each burst) grows to infinity with the network size N. Even if infected nodes maximize their infection capability to infect all neighbors and magnify the prevalence by a divergent factor, we demonstrate that the process is still localized and the spreading is restricted to a small subgraph, whose size divided by the whole network size N tends to zero.

In particular, our bursty SIS model is still an SIS model and is just the limiting case of the Weibullian SIS model with $\alpha \to \infty$ as introduced in Chapter 4. The bursty effect may lead to counterintuitive results. For example, in the epidemic process on a very large star graph, the infection probability of the hub node is much larger than those of the leaf node, when the process is just above the epidemic threshold. If the hub is infected just before a burst, the hub can infect all the leaf nodes and thus all nodes in the network are infected, which seems to lead to a non-zero prevalence (a global epidemic). However, even for the star graph, we will show that the prevalence just above threshold still converges to zero as the network size $N \to \infty$.

We denote the steady-state prevalence at time t^* after each burst by $y_\infty(\tilde{\tau}, t^*) \triangleq \frac{1}{N} \lim_{n\to\infty} \sum_{i=1}^{N} v_i(n/\beta + t^*)$ in the bursty SIS process with the normalized effective infection rate $\tilde{\tau} = \tau/\tau_c^{(B)} - 1$. The steady-state prevalence $y_\infty(\tilde{\tau}, t^*)$ is maximum just after each burst at $t^* = 0$, denoted by $y_\infty^+(\tilde{\tau}) \triangleq y_\infty(\tilde{\tau}, 0)$, and is minimum before each burst at $t^* \to 1/\beta$, denoted by $y_\infty^-(\tilde{\tau}) \triangleq \lim_{t^*\to 1/\beta} y_\infty(\tilde{\tau}, t^*)$. The ratio between the maximum and minimum steady-state prevalence is shown in Eq. (4.11) to be $y_\infty^+(\tilde{\tau})/y_\infty^-(\tilde{\tau}) \leq \lambda_1 + 1$ and equality is achieved when $\tilde{\tau} \downarrow 0$. Thus, for a network with a largest eigenvalue $\lambda_1 = O(N^\alpha)$ with $\alpha > 0$, $y_\infty^+(\tilde{\tau})/y_\infty^-(\tilde{\tau})$ diverges for small $\tilde{\tau}$ in the thermodynamic limit, which is the most unusual feature of the bursty dynamic compared to traditional studies. As shown in Fig. 5.2, the steady prevalence $y_\infty^+(\tilde{\tau})$ (blue curve) and $y_\infty^-(\tilde{\tau})$ (green curve) experience a phase transition at the threshold $\tilde{\tau} = 0$. Although the two curves approach each other from above to $\tau_c^{(B)}$, their ratio $y_\infty^+(\tilde{\tau})/y_\infty^-(\tilde{\tau})$ can diverge if $\lambda_1 \to \infty$ in the thermodynamic limit.

The maximum and the minimum steady-state prevalence $y_\infty^+(\tilde{\tau}) =$

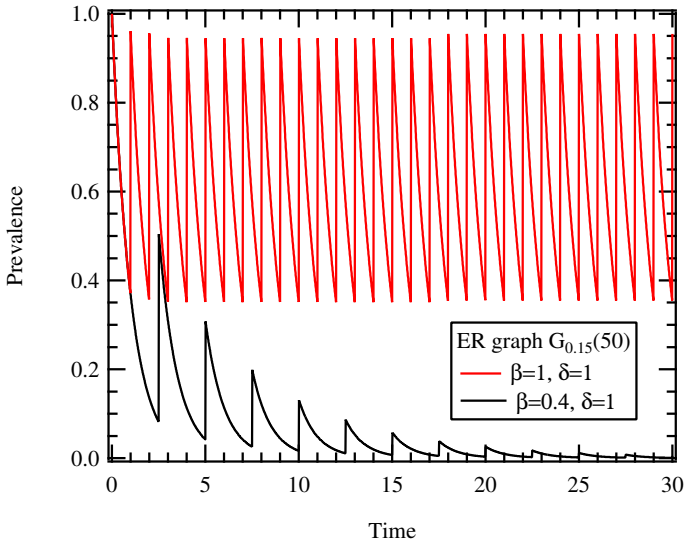

Figure 5.1: The bursty SIS prevalence on an Erdős-Rényi (ER) graph $G_{0.15}(50)$. The epidemic threshold is $\tau_c^{(B)} = \frac{1}{\ln(\lambda_1+1)} = 0.4437$. The red curve reflects the regime with the effective infection rate $\tau = 1 > \tau_c^{(B)}$, while the black curve represents the prevalence at $\tau = 0.4$ below the threshold.

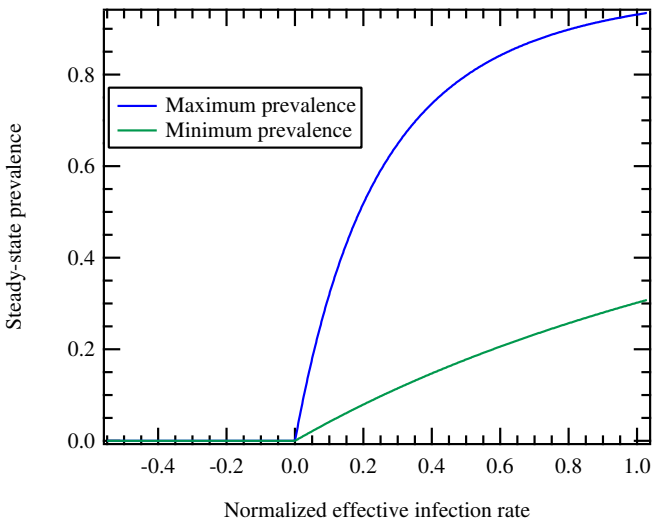

Figure 5.2: The phase transition of the bursty SIS model with the normalized effective infection rate $\tilde{\tau}$ on the same network. The upper blue curve and the lower green curve are the maximum and minimum steady-state prevalence, respectively. The steady-state prevalence changes periodically between the maximum and minimum.

$a_{\max}\tilde{\tau} + o(\tilde{\tau})$ and $y_{\infty}^{-}(\tilde{\tau}) = a_{\min}\tilde{\tau} + o(\tilde{\tau})$ just above threshold possess coefficients (see Theorem A.4.1 in Appendix A.4.1)

$$a_{\max} = \frac{2(\lambda_1 + 1)\ln(\lambda_1 + 1)}{\lambda_1} a \tag{5.2}$$

and $a_{\min} = a_{\max}/(\lambda_1 + 1)$, respectively. The coefficient a of the traditional SIS prevalence in (5.1) is only determined by the first- and third-order moments of the principal eigenvector x and the network size N, but the coefficients $a_{\max}$ and $a_{\min}$ are also related to the largest eigenvalue λ_1.

As mentioned, the bursts increase the prevalence by a factor of λ_1. For delocalized network with convergent maximum degree, we expect that the largest eigenvalue $\lambda_1 = O(1)$ because $\lambda_1 \leq \max_{\forall \text{link}(i,j)} \sqrt{d_i d_j}$ as shown in [90, p. 48]. Thus, the maximum and minimum prevalence are of the same order $O(1)$. There is always a non-zero average fraction of infected nodes just above the mean-field epidemic threshold in the thermodynamic limit.

Now we consider the localized networks. If the variance $\text{Var}[D] \to \infty$ as $N \to \infty$, then the largest eigenvalue $\lambda_1 \geq \sqrt{\text{Var}[D] + E^2[D]}$ diverges as shown in [90, p. 47]. Furthermore, a divergent maximum degree ensures the largest eigenvalue $\lambda_1 \to \infty$ as $N \to \infty$, since λ_1 of the whole network is larger than that of the star subgraph with a divergent hub [90, Eq. (3.23)]. In particular, the largest eigenvalue of a power-law network diverges in the thermodynamic limit [92]. The bursts magnify the traditional SIS coefficient a in (5.1) by a divergent factor $\ln(\lambda_1 + 1)$ as shown by Eq. (5.2), i.e. $a_{\max} = 2\ln(\lambda_1)a$. For the eigenvector localization as discussed in [186], where the eigenvector x is defined to be localized in a finite or infinite subnetwork, the coefficient a in (5.1) follows an decay as $O(N^{-\epsilon})$ for $\epsilon > 0$ and the maximum coefficient $a_{\max}$ in (5.2) will also converge to zero as $a_{\max} = O(N^{-\epsilon}\ln N)$ since $\ln \lambda_1 < \ln N$. Although the bursts allow the infected nodes to infect all their healthy neighbors to reach as many nodes as possible in the network, the bursts cannot transform a zero prevalence to a non-zero prevalence in the thermodynamic limit.

5.4 Numerical and simulation results

In this section, we evaluate our conclusion in synthetic and real networks.

5.4.1 Numerical results under the mean-field theory

The first case is the delocalized networks. In regular graphs with average degree d, the largest eigenvalue $\lambda_1 = d$ and the coefficients a_{max} and a_{min} are constant, only depending on degree d as explained in Appendix A.4.3. Figure 5.3 shows the results of the ER graphs with average degree $d_{\text{av}} = 8$, and both the maximum and minimum coefficients a_{max} and a_{min} are in the order of $O(1)$ and independent of the network size N.

For localized networks with divergent largest eigenvalue λ_1, the ratio between the maximum and minimum prevalence $\lim_{\tilde{\tau}\downarrow 0} y^+(\tilde{\tau})/y^-(\tilde{\tau}) \to \infty$ in the thermodynamic limit. We first consider star graphs as already mentioned. We may verify (see Appendix A.4.4) that the coefficients of star graphs follow $a_{\text{max}} = O(N^{-0.5} \ln N)$ and $a_{\text{min}} = O(N^{-1} \ln N)$. Although the average number of infected nodes both before and after each burst diverge, the maximum and minimum prevalence converges to zero as $N \to \infty$. We also generate connected scale-free networks with different power-law exponents γ and average degree $d_{av} = 8$ using the method introduced by Goh *et al.* [188]. When generating the scale-free networks, we only preserve the largest connected component, because the original method of Goh *et al.* does not guarantee a connected network. Figure 5.4 shows that the coefficient a_{max} of power-law networks with different exponent γ decays with the network size N. Furthermore, we consider networks with exponential degree distribution and use the network generating method in [74]. Initially, there are only m nodes in the network, and each step a new node arrives. The new node is randomly connected to m nodes of the current network (without preferential attachment as in the Barabási-Albert model [13]). The case $m = 1$ introduced in [74] generates a uniform recursive tree [88, 16.2.2]. Following a same derivation as in [74], the degree distribution of the network is $\Pr[D = k] = \frac{1}{1+m}(1 + 1/m)^{-k+m}$ for a network with average degree $d_{av} = 2m$ in the thermodynamic limit. Figure 5.5 shows the maximum coefficient a_{max} of exponential networks with $m = 1, 2, 4$, which decays with network size N.

For the synthetic networks, we can evaluate their near-threshold behavior by generating those networks with different size and check their order with the network size N. However, the size of a real network is fixed and the value of the coefficients a_{max} and a_{min} provide no information about the order of magnitude as a function of the network size N. To obtain insights from the value of a_{max} in real networks, we generate random synthetic networks with a similar size, average degree, and degree distribution for each real network and compare the coefficients a_{max} of the synthetic networks with those of the real

networks. For most real networks, the degree distributions approximately follow a power law[5] or exponential distribution. Thus, we can compare those real networks with the synthetic power-law and exponential networks mentioned above. Figure 5.6 shows the value of the coefficient $a_{\max}$ of real networks and corresponding synthetic networks, which are described in detail in the supplementary information. The value of the coefficients $a_{\max}$ are similar in synthetic and real networks, especially for large networks. Thus, we conjecture that the near-threshold behavior of bursts is similar in real and synthetic networks.

5.4.2 Simulations

We emphasize that the exact coefficient $a_{\max}$ is hard to obtain by simulations due to several reasons: a) The SIS process on finite-size networks has no sharp phase transition; b) Around the mean-field epidemic threshold, most realizations of the simulation die out (entering the absorbing all-healthy state) in a relatively short time. The time when the process is in the metastable state is hard to determine; c) The prevalence $y_\infty^+(\tilde{\tau})$ and the normalized effective infection rate $\tilde{\tau} = \tau/\tau_c^{(B)} - 1$ are small just above the mean-field threshold, and the numerical error of the exact coefficient $y_\infty^+(\tilde{\tau})/\tilde{\tau}$ can be large (since $\tilde{\tau} \approx 0$). Thus, only an approximation of the coefficient $a_{\max}$ can be obtained by simulations.

In our simulations of the bursty SIS process, all nodes are infected at time $t = 0$ to prevent early die-out [26]. If a node is infected at time t, then the infected node will be cured at time $t + T$ where T is an exponential random variable with mean $1/\delta$ and all its neighbors will be infected at time $t + k/\beta$ for $k = 1, 2, \ldots$, if $T > k/\beta$. Each realization of the bursty SIS process runs for 50 time units (simulations stop at $t = 50$) which are long enough under our setting and 10^5 realizations are simulated for each network. During the simulation of the bursty SIS process, the number of infected nodes is recorded every 0.01 time unit for each realization and the prevalence is calculated by averaging all realizations. The coefficient $a_{\max}$ is calculated by dividing the last local maximum of the recorded prevalence by $\tilde{\tau}$.

The simulation result on ER random graphs is shown in Fig. 5.3 for $\tilde{\tau} = 0.0001$ and curing rate $\delta = 4$. The results on power-law networks is shown in Fig. 5.4 for $\tilde{\tau} = 0.1$ and $\delta = 2$. We also perform the simulations

[5] Although there are debates that power-law networks are rare [80, 189, 81], the degree distribution of most real networks is linear in a log-log plot for several orders of magnitude, and then we can use synthetic power-law random graphs to approximate those real networks.

on exponential networks as shown in Fig. 5.5, for $\tilde{\tau} = 0.1$ with $\delta = 1$ for $m = 1, 2$ and $\delta = 2$ for $m = 4$. The different settings of parameters $\tilde{\tau}$ and δ are based on the relaxation time of the process, i.e. the time that the prevalence curve approaches zero visually. In the cases of power-law and the exponential graphs, most of the realizations die out and the prevalence is calculated by averaging the realizations which do not die out at $t = 45$. In the power-law and the exponential graphs, the simulation results are amazingly consistent with the mean-field theoretical results even though correlations of the infection state between neighbors are omitted in the mean-field analysis. In the ER graphs, the mean-field approximation does not perform well because the correlations play a role in sparse networks with homogeneous degree distribution [24]. However, the variation of the simulated coefficient $a_{\max}$ with the network size N agrees with the mean-field results: Fig. 5.3 indicates delocalization while Fig. 5.4 and Fig. 5.5 indicate localization of the bursty SIS process.

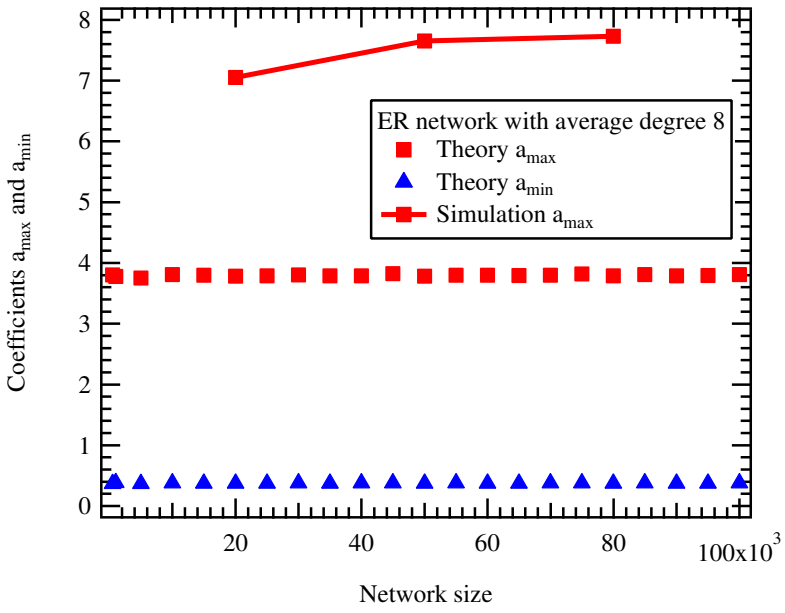

Figure 5.3: The coefficient $a_{\max}$ and $a_{\min}$ of ER networks.

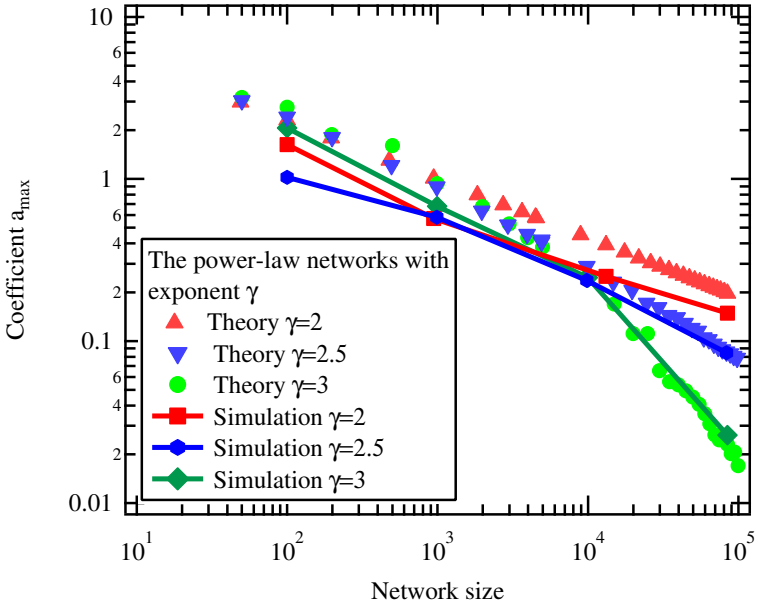

Figure 5.4: The coefficients $a_{\max}$ of networks with power-law degree distribution converge to zero with network size N.

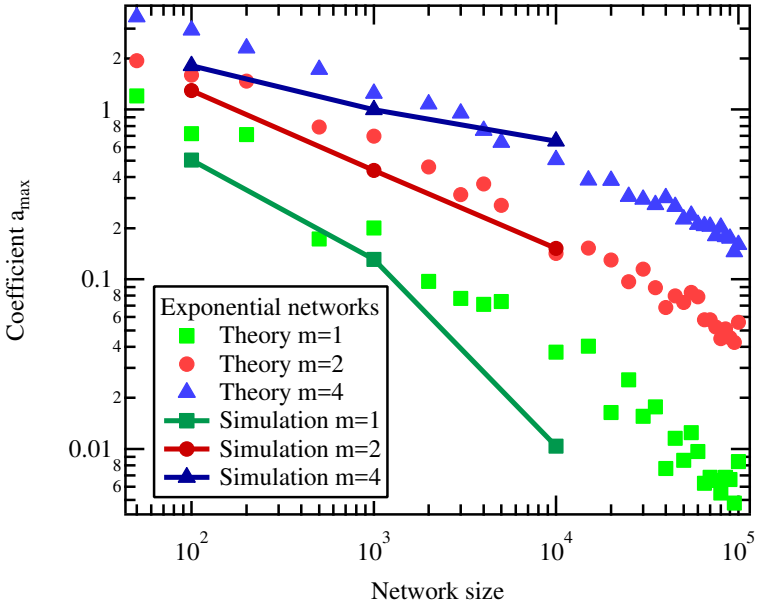

Figure 5.5: The coefficient $a_{\max}$ of networks with exponential degree distributions.

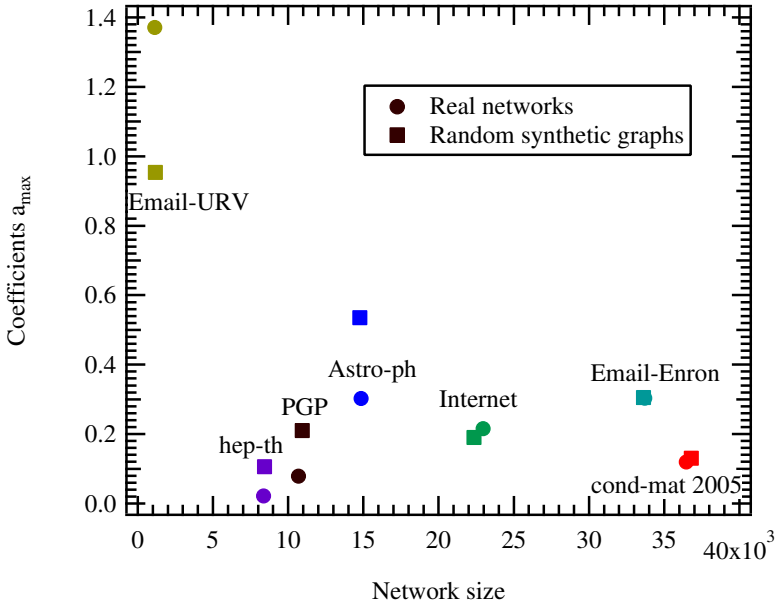

Figure 5.6: The coefficient $a_{\max}$ of some well studied real networks: Email-URV [190], hep-th [33], PGP [191], astro-ph [33], Internet [192], Email-Enron [193], and cond-mat 2005 [33].

5.5 Conclusion

In this chapter, we study the localization of the SIS process on networks. We specifically study a bursty SIS model which possesses a non-constant steady-state prevalence. In the bursty SIS model, the infected nodes can infect all healthy neighbors periodically to reach as many nodes as possible, and the prevalence is magnified by a divergent factor equal to the largest eigenvalue λ_1 in the thermodynamic limit. We show that the spreading process is still localized even if the bursty mechanism is applied, and our result introduces an open problem: are there any spreading dynamics leading to a delocalized spreading on networks with localized principal eigenvectors? If there exists such a case, then our analysis shows that the infection dynamic with a Poisson curing process must magnify the near-threshold prevalence $y_\infty(\tilde{\tau})$ of the traditional SIS model by a factor in the order of $O(N^z)$ for some value of $z \in (0, 1)$.

6

Pulse Strategy for Suppressing Spreading

In previous modelling efforts to understand the spreading process on networks, each node can infect its neighbors and cure spontaneously, and the curing is traditionally assumed to occur uniformly over time. This traditional curing is not optimal in terms of the trade-off between the effectiveness and cost. A pulse immunization/curing strategy is more efficient and broadly applied to suppress spreading process. We analyze the pulse curing strategy on networks with the Susceptible-Infected (SI) process. We analytically compute the mean-field epidemic threshold τ_c^p of the pulse SI model and show that $\tau_c^p = \frac{1}{\lambda_1} \ln \frac{1}{1-p}$, where λ_1 and p are the largest eigenvalue of the adjacency matrix of the contact graph and the fraction of nodes covered by each curing, respectively. These analytical results agree with simulations. Compared to the asynchronous curing process in the extensively studied Markovian SIS process, we show that the pulse curing strategy saves about 36.8%, i.e. $p \approx 0.632$, of the number of curing operations invariant to the network structure. Our results may help policymakers to design optimal containment strategies and minimize the controlling cost.[1]

6.1 Background

VIRAL spreading processes cause enormous losses of life. Due to the pandemic influenza A H1N1, 18500 laboratory-confirmed deaths are reported, while 284500 deaths are estimated during the period 2009.04 to 2010.08 [195]. Cyber-criminals earned around $\$100$ million per year by spreading an exploit kit, *Angler*, in computer systems [196]. A recent study

[1] This chapter is based on [194].

shows that false news spreads faster and more broadly than true news online [197]. The suppression of spreading processes is thus necessary in many circumstances, but consumes resources, e.g. budget in disease control or computational resources in detecting computer viruses. Based on the data from the World Health Organization, around 19.9 million children under the age of one still cannot receive the basic diphtheria-tetanus-pertussis (DTP3) vaccine and the coverage level of DTP3 for infants is only about 85% in 2017. Cisco reported [196] that 83% of the Internet of Things devices are not patched to be immunized against cyber-attacks.

Suppressing spreading requires a strategic design to balance between the cost and performance. A straightforward strategy is the uniform, asynchronous strategy: each infected individual can be cured uniformly over time as a Poisson process and thus the curing is asynchronous among infected individuals. This strategy is weak in preventing reinfections between direct neighbors because an cured individual can still have an infected neighbor. A pulse/synchronous strategy, where two direct neighbors have a high probability to be cured at the same time, is more efficient as shown in Fig. 6.1. The pulse strategy was first proposed to control the epidemic of measles [198] by periodically and synchronously vaccinating several age cohorts instead of uniformly and asynchronously vaccinating each individual at certain ages [199, 200]. In 1995, India introduced the National Immunization Days, which is a pulse strategy, to control the spread of polio [201]. Compared to the uniform, asynchronous strategy, the pulse strategy shows a better performance [202].

Furthermore, spreading processes are also focal in network science, because the underlying contact graph influences the spreading process nontrivially. For example, the epidemic threshold, which is determined by the network structure, of scale-free networks converges to zero with the network size under the mean-field approximation [12, 92, 102, 108, 184]. The spreading processes studied on networks are generally Markovian, which means that the infection and curing events occur both uniformly over time [88]. As mentioned earlier, the pulse strategy reduces the reinfections between neighboring nodes. If the curing occurs for all nodes at the same time, then no reinfection happens and the disease is immediately eradicated. If the curing only covers a fraction p of the whole population, synchronous curing with the pulse strategy still eliminates a substantial part of reinfections between neighbors and thus leads to better performance compared to a uniform, asynchronous curing strategy. Thus, one may wonder how efficient the pulse strategy is. The most reasonable way to quantify the effectiveness of the pulse strategy lies in assessing the reduction of the number of curing operations by using the asyn-

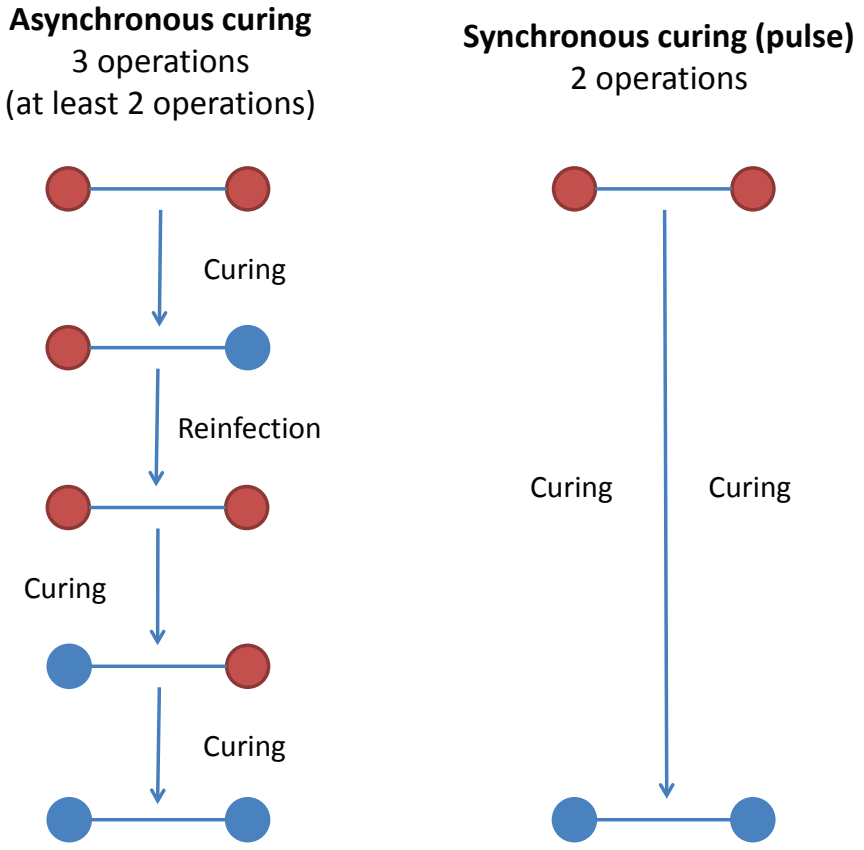

Figure 6.1: An example of a network with only two nodes. The asynchronous curing conducts at least two curing operations because every reinfection adds one curing operation while the synchronous curing conducts only two curing operations to eliminate the spreading.

chronous strategy as a benchmark. In the following, we consider the most basic spreading model on networks, i.e., the Susceptible-Infected (SI) process, and evaluate the pulse strategy performed on the SI model. Here, we refer to the curing as a strategy, because we are focusing on curing actions that can be performed in synchronous/pulse or asynchronous manner by the public health department or cybersecurity team. In contrast and beyond our scope, individuals may be spontaneously cured by the immune system during an epidemic outbreak, which is essentially an asynchronous curing.

6.2 The model: the SI process with pulse curing

In the networked spreading process, each node in the network is either infected or susceptible (healthy). Each infected node can infect each healthy neighbor by a Poisson process with rate β. We assume that each node is cured with rate δ. Thus, for the pulse curing strategy, the curing happens every $1/\delta$ time units, i.e. the nodes can only be synchronously cured at time k/δ for $k = 1, 2, \ldots$. The curing has a successful probability p turning an infected node into a healthy one. Equivalently, each node can be cured certainly, but only a fraction p of nodes are randomly chosen to be cured. We define the effective infection rate $\tau \triangleq \beta/\delta$.

The difference between the above pulse curing SI model and the extensively studied Markovian Susceptible-Infected-Susceptible (SIS) model [11] is that each node in the Markovian SIS model is cured by a Poisson process with rate δ and $p = 1$, which represents an asynchronous curing strategy. In the Markovian SIS process on networks, there exists an epidemic threshold [102, 108] under the N-Intertwined mean-field approximation $\tau_c^{(1)} = \frac{1}{\lambda_1}$ where λ_1 is the largest eigenvalue of the adjacency matrix of the network. If $\tau > \tau_c^{(1)}$, then the process is in an endemic phase in the steady state, but if $\tau < \tau_c^{(1)}$, then the process converges to the all-healthy state. In the pulse curing strategy, limited resources or some other complications may lead to a partial coverage specified by a fraction $p < 1$. If $p = 1$, then synchronous curing destroys the spreading immediately. The average numbers of curing operations in the asynchronous Poisson curing and the pulse curing are δ and δp, respectively, for each node during one unit of time. In the following, we analyze the pulse curing effect on epidemic processes on networks under the mean-field theory to derive the epidemic threshold. Our main finding is that when $p = 1 - 1/\mathrm{e} \approx 0.632$, the pulse curing is equally effective to the Poisson curing process with the same curing rate δ.

6.2.1 Mean-field analysis and the epidemic threshold

We represent the time t in the form of $t = k/\delta + t^*$, where $t^* \in [0, 1/\delta)$. For $t^* \neq 0$, only infection happens and the mean-field equation of node i is

$$\frac{dv_i(k/\delta + t^*)}{dt^*} = \beta \left[1 - v_i(k/\delta + t^*)\right] \sum_{j=1}^{N} a_{ij} v_j(k/\delta + t^*) \tag{6.1}$$

where $v_i(k/\delta + t^*)$ is the probability that node i is infected at time $t = k/\delta + t^*$ and $a_{ij} \in \{0, 1\}$ is the element of adjacency matrix of the network with N nodes. The probability $v_i(k/\delta + t^*)$ is discontinuous at $t^* = 0$ for all k when curing happens: $\lim_{t^* \to 0} v_i(k/\delta + t^*) = v_i(k/\delta)$ and $\lim_{t^* \to 1/\delta} v_i(k/\delta + t^*) \neq v_i((k+1)/\delta)$. Equation (6.1) is a mean-field approximation, because we omit the correlation of the infection state between neighbors just as in the Markovian SIS process [110]. Since the curing probability of each node at k/δ is p, the pulse curing process is governed by the following equation,

$$v_i\left(\frac{k+1}{\delta}\right) = (1-p) \lim_{t^* \to 1/\delta} v_i\left(\frac{k}{\delta} + t^*\right) \tag{6.2}$$

In our previous work [156], we introduced the bursty SIS model, where the infection happens periodically with rate β and the curing is a Poisson process. In the bursty SIS model, the relationship between the infection probability of each node at the start $t^* = 0$ and the end $t^* \to 1/\beta$ of the same time interval is explicitly known as an exponentially decreasing function. In pulse curing, the relationship between $v_i(k/\delta)$ and $\lim_{t^* \to 1/\delta} v_i(k/\delta + t^*)$ is described by (6.1), which does not have an explicit solution for general networks[2]. However, since we only care about the regime where $v_i(k/\delta + t^*) \to 0$ to derive the epidemic threshold, we can first linearize Eq. (6.1) around $v_i(k/\delta + t^*) = 0$ for all i and obtain

$$\frac{d\mathbf{v}(k/\delta + t^*)}{dt^*} = \beta A \mathbf{v}(k/\delta + t^*) \tag{6.3}$$

where the infection probability vector $\mathbf{v}(k/\delta + t^*) \triangleq [v_1(k/\delta + t^*), \ldots, v_N(k/\delta + t^*)]^T$. The general solution [88, p. 209] of (6.3) is $\mathbf{v}(k/\delta + t^*) = \mathrm{e}^{\beta A t^*} C$ where $C = \mathbf{v}(k/\delta)$ is the initial value vector at $t^* = 0$.

[2]Only for the regular graph when the initial condition of each node is identical, there is an explicit solution for (6.1). One may verify for the d-regular graph that $v_i(k/\delta) = (1-p) - e^{-d\tau}$. Let $v_i(k/\delta) = 0$ and the threshold is $\frac{1}{d} \ln \frac{1}{1-p}$ which is consistent with (6.6)

Thus, the solution of Eq. (6.3) evaluated at $t^* \to 1/\delta$ is

$$\lim_{t^* \to 1/\delta} \mathbf{v}(k/\delta + t^*) = \mathrm{e}^{\tau A}\mathbf{v}(k/\delta) \tag{6.4}$$

Substituting (6.4) into the curing equation (6.2) yields

$$\mathbf{v}\left(\frac{k+1}{\delta}\right) = (1-p)\mathrm{e}^{\tau A}\mathbf{v}\left(\frac{k}{\delta}\right) \tag{6.5}$$

When the largest eigenvalue of $(1-p)\mathrm{e}^{\tau A}$, which is $(1-p)e^{\tau\lambda_1}$, is smaller than one, eq. (6.5) shows that the infection probability $\mathbf{v}\left(\frac{k}{\delta}\right)$ converges to zero in the long run. Thus, for $(1-p)e^{\tau\lambda_1} = 1$, we obtain the epidemic threshold

$$\tau_c^p \triangleq \frac{1}{\lambda_1}\ln\frac{1}{1-p} \tag{6.6}$$

If $\tau > \tau_c^p$, then the spreading can persist in the network, which is the endemic phase. If $\tau < \tau_c^p$, then the spreading disappears in the long run, which is the all-healthy phase.

The Markovian SIS process with a Poisson curing process has a mean-field epidemic threshold $\frac{1}{\lambda_1}$. When $\ln\frac{1}{1-p} = 1$ or $p = 1 - 1/\mathrm{e} \approx 0.632$, the pulse curing is equivalent to the Poisson curing process in the traditional SIS model on any graph. Thus, to eliminate the spreading, the pulse strategy only consumes 63.2% of the number of curing operations of the asynchronous strategy, since the curing rates δ of the two strategies are equal. In the next section, two typical examples show that even above the epidemic threshold, the two strategies are comparable, if $p = 0.632$.

6.2.2 Simulation: above the epidemic threshold

In Fig. 6.2, we show the prevalence, which is the average fraction of the infected nodes, of the Markovian SIS model and the pulse curing model with $p = 0.632$, on a BA graph and an ER graph. The effective infection rates τ are above the epidemic thresholds $1/\lambda_1$. The prevalence of the Markovian SIS model is exactly centered in the middle of the prevalence generated by the pulse curing SI model. Figure 6.2 indicates that the two curing processes are equivalent to some extent at $p = 0.632$, even above the epidemic threshold.

6.2.3 Phase diagram

Figure 6.3 shows the phase diagram of the pulse curing strategy with the mean-field epidemic threshold calculated by (6.6). For small coverage p, the thresh-

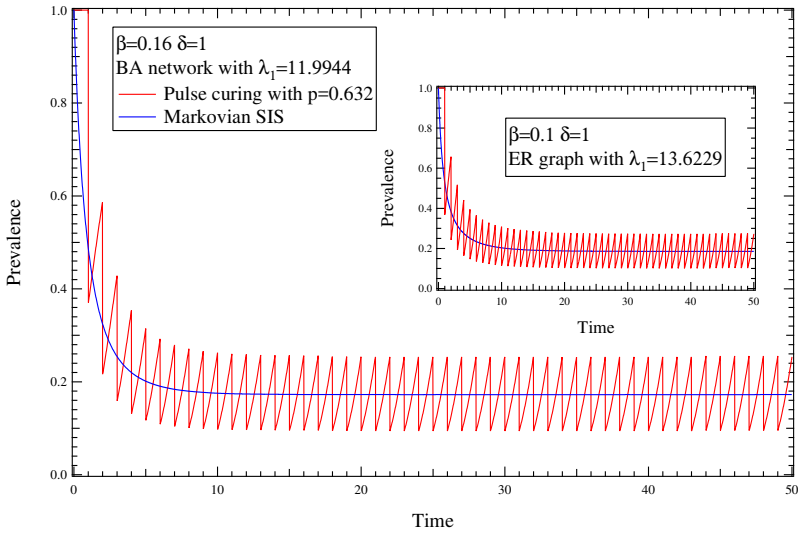

Figure 6.2: The prevalence of the Markovian SIS model and the pulse curing model obtained by averaging 10^5 simulated realizations. The simulation is performed on a $N = 500$-node network generated by the Barabási-Albert model and a $N = 500$-node ER graph. The curing probability is set to be $p = 0.632$ for the pulse strategy.

old τ_c^p increases slowly with p; While for large p, there is an increased effectiveness of p in the pulse strategy.

For a spreading process in the endemic phase, one can tune both the curing rate δ and the curing coverage p to move the process from the endemic phase to the all-healthy phase. Figure 6.3 shows that there are many different ways to achieve this. However, the optimal way is just to increase the curing coverage p and to decrease the curing rate δ along the red curve. The argument is as follows. From (6.6), we have that $\delta = \lambda_1\beta/\ln[1/(1-p)]$ and thus $\delta p = \lambda_1\beta p/\ln[1/(1-p)]$, when $\tau = \tau_c^p$. The goal is to minimize the average number of curing operations δpN during one time unit, which asks to minimize $p/\ln[1/(1-p)]$. One may verify that $p/\ln[1/(1-p)]$ is monotonically decreasing with p in $(0,1)$ and thus increasing p along the red curve in the phase diagram is the optimal way of choosing δ and p to suppress spreading. The result is reasonable, because a large p can probably shut down the spreading within a few curing pulses. In real scenarios, the coverage p may be restricted and thus choosing the maximum possible p and a corresponding δ is an option.

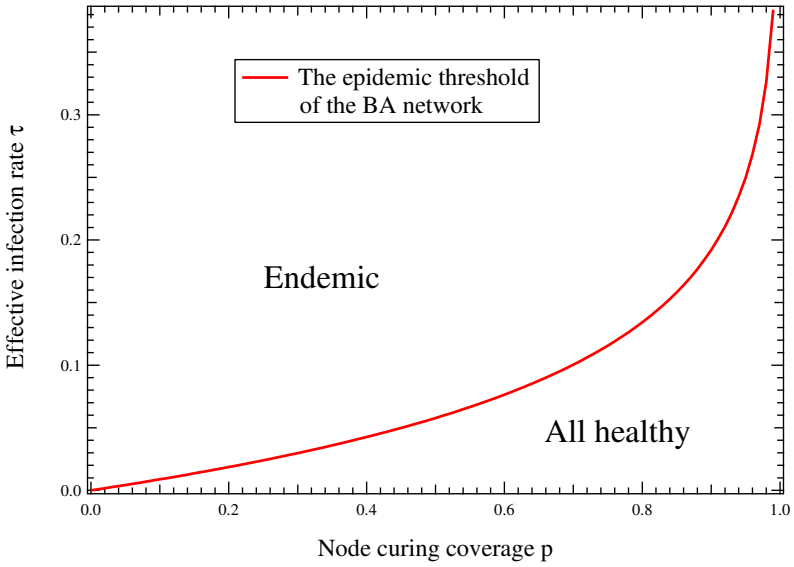

Figure 6.3: The phase diagram of the BA network calculated by Eq. (6.6).

6.3 Conclusion

We quantified the effect of the pulse strategy for suppressing spreading processes on networks. To achieve an equivalent effect, the pulse strategy consumes 63.2% of the total number of curing operations, required by the uniform and asynchronous strategy, e.g. a Poisson curing process. This reduction of cost does not depend on the underlying contact graph in the mean-field approximation. Our results may help the agencies, e.g. disease control centers or computer security teams, to make policies or allocate resources.

7 Conclusion

"All models are wrong, but some are useful."

—George Box, 1978.

BY presenting new knowledge about the basic SIS model, we try to reveal the underlying mechanism of spreading processes on networks. We study both Markovian and non-Markovian models on unweighted, undirected networks. The analysis is under mean-field approximations where the correlations between neighbors are omitted. The analytic results are compared with direct simulations of the exact processes from the definitions. Our work contributes both to modelling spreading phenomena and understanding complex systems.

7.1 Main contributions

For the Markovian SIS model on networks, we show that the mean-field approximations can be applied to calculate the correlation of the infection state of the same node at different time points both in the transient and steady state. The autocorrelation is a function of the mean-field infection probabilities of nodes. Under the mean-field approximation, the infection transition of each node can be denoted by a two-state Markov process with a time-dependent transition rate in the transient state and with a time-invariant rate in the steady state. In the steady state, the calculation of the infection autocorrelation is fairly straight forward. In the transient state, the autocorrelation can be obtained by solving the time-inhomogeneous Kolmogorov forward equation of the two-state Markov process using the Magnus expansion. We find that the

autocorrelation of the infection state is independent of the curing rate in regular graphs. The infection and curing rate can be estimated given a sequential observation of the infection state in the steady state. For the estimation of the curing rate, the network structure is not needed.

To understand non-Markovian spreading processes, we substitute the Poissonian infection process by a renewal process whose inter-arrival time follows a Weibull distribution. This Weibullian SIS process is an extension of the Markovian SIS process since the Poisson process is a special case of the Weibull renewal process. We investigate the limiting case where the Weibull probability density function is a Dirac delta function. In the limiting case, the infection becomes synchronized and periodical at all nodes and each infection infects all healthy neighbors if the initially infected nodes are seeded at the same time. The mean-field epidemic threshold of the limiting Weibullian SIS process is shown to be $\tau_c^{(B)} = \frac{1}{\ln(\lambda_1+1)}$. By simulation, we show $\tau_c^{(B)}$ is the largest epidemic threshold among all the shapes of the Weibull distributed infection time. Moreover, the Gamma and lognormal time distributions in the renewal infection process lead to similar behavior of the SIS process on networks.

We further study the limiting case of the Weibullian SIS model just above the epidemic threshold. In the near-threshold steady state, the infection probability of each node changes periodically and the ratio between the maximum and minimum infection probabilities reaches its largest value $1+\lambda_1$. We calculate the steady maximum prevalence as a function of the normalized effective infection rate. The coefficient of the first-order term in its Taylor expansion depends not only on the principal eigenvector of the adjacency matrix as the Markovian SIS model but also on the principal eigenvalue. Our analysis indicates that if the Markovian SIS process is localized in a network, then the limiting Weibullian SIS process is also localized on the same network. Although near the threshold, the ratio between the maximum and the minimum prevalence $\lambda_1 + 1$, which seems to be a maximum amplification of the prevalence, diverges in the thermodynamic limit, the maximum prevalence still tends to zero and the spreading process is localized. The results imply that the localization of the spreading process is only determined by the network structure but not the infection pattern.

Finally, we quantify the efficiency of the pulse curing strategy by comparing the pulse curing with the uniform curing. The Poissonian curing process is substituted by a periodical synchronized curing process that performs on a fraction $0 < p < 1$ of all nodes. The epidemic threshold is shown to be

$\tau_c^p = \frac{1}{\lambda_1} \ln \frac{1}{1-p}$. Let τ_c^p equal to the threshold $\frac{1}{\lambda_1}$ of the Markovian SIS model, which represents adopting a uniform curing strategy and then $p = 0.632$. Thus, the pulse strategy is equivalent to the uniform strategy if only 63.2% curing operations are performed. By simulation, the pulse curing is also equivalent to the uniform one even above the threshold in the sense that the time-dependent prevalence of the Markovian SIS process is exactly located at the center of the oscillating prevalence of the SIS model adopting the pulse strategy. The optimal planning of suppressing spreading is discussed based on the mean-field analysis.

7.2 Future works

The SIS model is simple by definition, but the analysis is complicated and the model is still not understood fully. As the Markovian SIS process is a linear system, there is no sharp phase transition for finite-size networks. For general infinite large networks, it is yet known whether the epidemic threshold predicted by mean-field theories equals to the exact epidemic threshold or not. What parameters of the network does the accuracy of mean-field theories relate to?

In Section 5.4.2, the simulation results in Fig. 5.4 show that NIMFA is very accurate in some networks if we remove the die-out realizations in the simulation. The die-out in Section 5.4.2 is mainly due to the small infection rate: Since there is no sharp epidemic threshold in the exact process, the epidemic process can die out with a high probability just above the mean-field threshold. Also, we have discussed the inaccuracy due to the initial die-out in Section 2.4 and find similarly that the mean-field approximation is also accurate comparing to the simulations where the die-out realizations are removed. The difference here is that the die-out in Section 2.4 is caused by the small number of initially infected nodes and the infection rate is far from the epidemic threshold. The assumption of independence between neighbors in NIMFA leads to a non-zero steady state which is globally stable in the sense that the SIS process always evolves to this non-zero steady state from a non-zero initial condition. In other words, there is no die-out in NIMFA above the NIMFA threshold and the following proposition needs to be studied: The only effect of omitting the correlation between neighbors is removing die-out in the exact process or equivalently, the die-out is caused by the correlation between neighbors. If the proposition is true, then in the thermodynamic limit where no die-out exists in the exact process, it might be true that the mean-field is

equivalent to the exact process. Of course, the above argument does not consider the sparse homogeneous networks, such as the cycle graphs, on which the mean-field is not accurate even removing the die-out realizations.

Another question is about the Griffiths phase and localization of SIS processes. In Chapter 5, we have introduced the localization phenomena of the SIS processes. Just above the mean-field thresholds, the prevalence tends to zero in the thermodynamic limit. Other studies [19, 15] claim that just above the mean-field threshold is the Griffiths phase, which means that the prevalence is a non-analytic smooth function above the threshold. However, a rigorous mathematical analysis is still absent for general networks.

A Appendices

A.1 Appendix of Chapter 2

A.1.1 The die-out probability

The expectation of the fraction of infected node $\tilde{I}(t)$ of an SIS process survived until t, i.e. $I(t) \neq 0$, in a network with N nodes is

$$E[\tilde{I}(t)] = \sum_{i=1}^{N} \frac{i}{N} \Pr[\tilde{I}(t) = \frac{i}{N}] = \sum_{i=1}^{N} \frac{i}{N} \Pr\left[I(t) = \frac{i}{N} \middle| I(t) \neq 0 \right]$$

With the definition of the conditional probability,

$$\begin{aligned} \Pr\left[I(t) = \frac{i}{N} \middle| I(t) \neq 0 \right] &= \frac{\Pr\left[\left\{ I(t) = \frac{i}{N} \right\} \cap \{ I(t) \neq 0 \} \right]}{\Pr[I(t) \neq 0]} \\ &= \frac{\Pr\left[I(t) = \frac{i}{N} \right]}{\Pr\left[I(t) \neq 0 \right]} \text{ provided } i > 0 \end{aligned}$$

we have

$$E[\tilde{I}(t)] = \frac{1}{\Pr[I(t) \neq 0]} \sum_{i=0}^{N} \frac{i}{N} \Pr\left[I(t) = \frac{i}{N} \right] = \frac{E[I(t)]}{\Pr[I(t) \neq 0]}$$

Since $\Pr[I(t) \neq 0] = 1 - \Pr[I(t) = 0]$, the prevalence can be written as

$$y(t) = \tilde{y}(t) \left(1 - \Pr[I(t) = 0] \right) \tag{A.1}$$

where $\tilde{y}(t) = E[\tilde{I}(t)]$.

A.1.2 The bounds of the spreading time through a path

Due to the absorbing state of the Markovian SIS process, the expected extinction time of a birth-death process from state n to the absorbing state 0 is [203],

$$t_n = \sum_{m=1}^{n} \left[\frac{1}{\mu_m} + \left(\prod_{i=1}^{m-1} \frac{\mu_i}{\lambda_i} \right) \sum_{j=m+1}^{N} \frac{1}{\mu_j} \prod_{k=1}^{j-1} \frac{\lambda_k}{\mu_k} \right] \tag{A.2}$$

where λ_i and μ_i are birth and death rate at state i, respectively.

The upper bound of the SIS spreading time through a path can be calculated by difference equations. In the upper bound model as mentioned above, we assume that the expected extinction time from state k is t_k. The process stays at state k initially, and then two possible state transitions can happen: $k \to L$ with probability $\frac{\delta}{\beta+\delta}$, and $k \to k-1$ with probability $\frac{\beta}{\beta+\delta}$. Thus, the extinction time from state k in this model equals to the sum of the sojourn time of state k and the average extinction time of t_{k-1} and t_L, which can be calculated by the law of total probability. The difference equations for $1 \leq k \leq L-1$ can be written as,

$$t_k = \frac{1}{\beta+\delta} + \frac{\beta}{\beta+\delta} t_{k-1} + \frac{\delta}{\beta+\delta} t_L \tag{A.3}$$

where $t_L = t_U$ is the upper bound of the spreading time through an L-length path. We have $t_0 = 0$, and a reflecting boundary at state L is $t_L - t_{L-1} = \frac{1}{\beta}$. Equation (A.3) also holds when $k = L$. We rewrite (A.3) by dividing $(\frac{\beta}{\beta+\delta})^k$ on both sides and rearrange the terms as,

$$\frac{t_k}{\left(\frac{\beta}{\beta+\delta}\right)^k} - \frac{t_{k-1}}{\left(\frac{\beta}{\beta+\delta}\right)^{k-1}} = \left(\frac{1}{\beta+\delta} + \frac{\delta}{\beta+\delta} t_L \right) \left(\frac{\beta}{\beta+\delta} \right)^{-k} \tag{A.4}$$

Let $A_k = \frac{t_k}{\left(\frac{\beta}{\beta+\delta}\right)^k} - \frac{t_{k-1}}{\left(\frac{\beta}{\beta+\delta}\right)^{k-1}}$. By summing over A_k for $k = 1, 2, \cdots, L$, we obtain the upper bound of the spreading time of the SIS process through an L-length path,

$$\begin{aligned} t_L &= \left(\frac{\beta}{\beta+\delta} \right)^L \left(\sum_{k=1}^{L} A_k \right) \\ &= \left(\frac{\beta}{\beta+\delta} \right)^L \left(\frac{1}{\beta+\delta} + \frac{\delta}{\beta+\delta} t_L \right) \sum_{k=1}^{L} \left(\frac{\beta}{\beta+\delta} \right)^{-k} \end{aligned} \tag{A.5}$$

The upper bound of the spreading time t_U as (2.18) can be derived by rearranging equation (A.5).

A.2 Appendix of Chapter 3

A.2.1 Autocorrelation of the infection state and the fraction of infected nodes in the steady state

When the effective infection rate $\tau > 1/\lambda_1$, the steady infection probability $v_{j\infty}$ can be obtained by solving $dv_j(t)/dt = 0$, i.e.,

$$-\delta v_{j\infty} + (1 - v_{j\infty})\tilde{\beta}_{j\infty} = 0 \tag{A.6}$$

where $\tilde{\beta}_{j\infty} = \lim_{t\to\infty} \tilde{\beta}_j(t)$ is time-invariant. Thus,

$$\tilde{\beta}_{j\infty} = \frac{\delta v_{j\infty}}{1 - v_{j\infty}} \tag{A.7}$$

and the steady infinitesimal generator of node j is

$$Q_{j\infty} \triangleq \lim_{t\to\infty} Q_j(t) = \begin{bmatrix} -\tilde{\beta}_{j\infty} & \tilde{\beta}_{j\infty} \\ \delta & -\delta \end{bmatrix}$$

In the steady state, the transition probability matrix of $V_{j\infty}(t)$ with time lag h is

$$P_{j\infty}(h) \triangleq \begin{bmatrix} \Pr[V_{j\infty}(t+h) = 0 | V_{j\infty}(t) = 0] & \Pr[V_{j\infty}(t+h) = 1 | V_{j\infty}(t) = 0] \\ \Pr[V_{j\infty}(t+h) = 0 | V_{j\infty}(t) = 1] & \Pr[V_{j\infty}(t+h) = 1 | V_{j\infty}(t) = 1] \end{bmatrix}$$

By solving the Kolmogorov forward equation $P'_{j\infty}(h) = P_{j\infty}(h)Q_{j\infty}$ given that $P_{j\infty}(0)$ is an identity matrix, we obtain

$$\begin{aligned} P_{j\infty}(h) &= \mathrm{e}^{Q_{j\infty}h} \\ &= U_j \mathrm{e}^{\Lambda_j} U_j^{-1} \end{aligned} \tag{A.8}$$

where U_j and Λ_j are the eigenvector matrix and the diagonal eigenvalue matrix

of $Q_{j\infty}h$, respectively. The term $Q_{j\infty}h$ can be diagonalized as

$$Q_{j\infty}h = \underbrace{\begin{bmatrix} -\frac{\tilde{\beta}_{j\infty}/\delta}{\sqrt{\left(\tilde{\beta}_{j\infty}/\delta\right)^2+1}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{\left(\tilde{\beta}_{j\infty}/\delta\right)^2+1}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{U_j} \underbrace{\begin{bmatrix} -(\tilde{\beta}_{j\infty}+\delta)h & 0 \\ 0 & 0 \end{bmatrix}}_{\Lambda_j}$$
$$\underbrace{\left(-\sqrt{2}\frac{\sqrt{\left(\tilde{\beta}_{j\infty}/\delta\right)^2+1}}{\tilde{\beta}_{j\infty}/\delta+1} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{\left(\tilde{\beta}_{j\infty}/\delta\right)^2+1}} & -\frac{\tilde{\beta}_{j\infty}/\delta}{\sqrt{\left(\tilde{\beta}_{j\infty}/\delta\right)^2+1}} \end{bmatrix} \right)}_{U_j^{-1}}$$

By substituting $e^{\Lambda_j} = \begin{bmatrix} \mathrm{e}^{-(\tilde{\beta}_{j\infty}+\delta)h} & 0 \\ 0 & \mathrm{e}^0 \end{bmatrix}$ and U_j into (A.8), we obtain the steady-state transition probability matrix

$$P_{j\infty}(h) = \begin{bmatrix} \frac{\tilde{\beta}_{j\infty}\mathrm{e}^{-(\tilde{\beta}_{j\infty}+\delta)h}+\delta}{\tilde{\beta}_{j\infty}+\delta} & \frac{-\tilde{\beta}_{j\infty}\mathrm{e}^{-(\tilde{\beta}_{j\infty}+\delta)h}+\tilde{\beta}_{j\infty}}{\tilde{\beta}_{j\infty}+\delta} \\ \frac{-\delta\mathrm{e}^{-(\tilde{\beta}_{j\infty}+\delta)h}+\delta}{\tilde{\beta}_{j\infty}+\delta} & \frac{\delta e^{-(\tilde{\beta}_{j\infty}+\delta)h}+\tilde{\beta}_{j\infty}}{\tilde{\beta}_{j\infty}+\delta} \end{bmatrix} \tag{A.9}$$

From (A.9), the joint expectation for $h \geq 0$ is

$$\begin{aligned} E[V_{j\infty}(t)V_{j\infty}(t+h)] &= \Pr[V_{j\infty}(t)=1, V_{j\infty}(t+h)=1] \\ &= \Pr[V_{j\infty}(t+h)=1|V_{j\infty}(t)=1]\Pr[V_{j\infty}(t)=1] \\ &= v_{j\infty}\frac{\delta e^{-(\tilde{\beta}_{j\infty}+\delta)h}+\tilde{\beta}_{j\infty}}{\tilde{\beta}_{j\infty}+\delta} \end{aligned} \tag{A.10}$$

By substituting $E[V_j(t)V_j(t+h)]$ from (A.10) and $\tilde{\beta}_{j\infty}$ from (A.7) into (3.2), we obtain (3.3).

The autocorrelation of the fraction of infected nodes $I_{j\infty}(t)$ is,

$$\begin{aligned} R_{I\infty}(h) &= \frac{E[I_\infty(t)I_\infty(t+h)] - E[I_\infty(t)]E[I_\infty(t+h)]}{\mathrm{Var}[I(t)]} \\ &= \frac{\sum_{j=1}^{N}\left(E[V_{j\infty}(t)V_{j\infty}(t+h)] - v_{j\infty}^2\right)}{\sum_{j=1}^{N}(v_{j\infty}-v_{j\infty}^2)} \end{aligned} \tag{A.11}$$

From (3.2), the term $E[V_{j\infty}(t)V_{j\infty}(t+h)]-v_{j\infty}^2 = R_{j\infty}(h)\left(v_j - v_j^2\right)$. Thus, we obtain (3.8).

A.2.2 The HMF autocorrelation in the steady state

HMF assumes the SIS process is running on an annealed network where nodes with the same degree are statistically equivalent [11] or a time-varying network with infinite rewiring rate [204]. On static networks, NIMFA performs better [145]. The HMF equation is

$$\frac{dw_d(t)}{dt} = -\delta w_d(t) + \beta(1 - w_d(t))d \sum_{k=1}^{N-1} f(k,d) w_k(t) \tag{A.12}$$

where $w_d(t)$ denotes the infection probability of the nodes with degree d, and $f(k,d)$ is the probability that an edge of a node with degree d connects to a node with degree k. The HMF threshold is $\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]}$, where D is the degree of a randomly selected node. We assume $\tilde{\beta}_d^{\text{HMF}}(t) = \beta d \sum_{k=1}^{N-1} f(k,d) w_k(t)$. In the steady state $dw_d(t)/dt = 0$ when $\tau > \tau_c^{\text{HMF}}$, the HMF infection probability $w_{d\infty} \triangleq \lim_{t\to\infty} w_d(t)$ follows

$$-\delta w_{d\infty} + (1 - w_{d\infty})\tilde{\beta}_{d\infty}^{\text{HMF}} = 0 \tag{A.13}$$

where $\tilde{\beta}_{d\infty}^{HMF} = \lim_{t\to\infty} \tilde{\beta}_d^{HMF}(t)$. Equation (A.13) has a same form with (A.6), and the derivation of the HMF autocorrelations is also similar. For example in the steady state, following the same derivation in Appendix A.2.1, the HMF autocorrelation of nodes with degree d is

$$R_{d\infty;\text{HMF}} = \mathrm{e}^{-\frac{\delta}{1-w_{d\infty}}h}$$

The HMF autocorrelation of the fraction of infected nodes is,

$$R_{I\infty;\text{HMF}}(h) = \frac{\sum_{d=1}^{N}[\Pr(D=d)]^2 R_{d\infty;\text{HMF}}(w_{d\infty} - w_{d\infty}^2)}{\sum_{d=1}^{N}[\Pr(D=d)]^2 (w_{d\infty} - w_{d\infty}^2)}$$

Here, we assume that the HMF fraction of infection nodes is $\sum_{d=1}^{N} \Pr(d) W_d(t)$, where $W_d(t)$ is the infection state of nodes with degree d. The state transition of nodes with the same degree is considered as coupled Markov processes and the infection states of nodes with the same degree are same. In the case of regular graphs, HMF and NIMFA are equivalent and then the two approximate autocorrelations are equal.

A.2.3 The Magnus expansion for time-inhomogeneous Markov processes

In this section, we shortly introduce the Magnus expansion, and then apply the Magnus expansion to NIMFA to derive the necessary results used in analyzing the NIMFA transition matrix.

In a time-inhomogeneous Markov process with D states, the $D \times D$ probability transition matrix $P(s,t)$ from time s to time t follows the Kolmogorov forward equation

$$\frac{dP(s,t)}{dt} = P(s,t)Q(t) \tag{A.14}$$

where $Q(t)$ is the $D \times D$ time-dependent infinitesimal generator. Stroock [205, p. 164] analyzes Eq. (A.14) by dividing time into smaller subintervals with length $1/n$. In each subinterval, the infinitesimal generator $Q(t)$ is assumed to be constant and an approximate transition matrix $P^{[n]}(s,t)$ can be obtained. The transition matrix $P^{[n]}(s,t)$ converges to the unique solution $P(s,t)$ when $n \to \infty$, which follows the time-inhomogeneous Chapman-Kolmogorov equation,

$$P(s,t) = P(s,r)P(r,t) \tag{A.15}$$

for $s \leq r \leq t$. However, n is always finite for the practical calculation of $P(s,t)$, and Stroock's method do not give a hint on the accuracy of the calculation.

More generally, the linear Eq. (A.14) always [147, p. 166] has a unique solution in form $P(s,t) = \exp(\Omega(s,t))$, where $\Omega(s,t)$ is an $D \times D$ matrix. If the commutative property $Q(t_1)Q(t_2) = Q(t_2)Q(t_1)$ holds for any $t_1, t_2 \in [s,t]$, then $\Omega(s,t) = \int_s^t Q(u)du$ and Eq. (A.14) has a closed form solution $P(s,t) = \exp(\int_s^t Q(u)du)$. However, the commutative property does not necessarily hold in most, if not all, time-inhomogeneous Markov processes.

Equation (A.14) can be analyzed using the Magnus expansion [146] when $Q(t)$ is not commutative. In a small time interval $t \in [s, s+T]$ such that [147, Theorem 9]

$$\int_s^{s+T} \|Q(t)\|dt < \pi \tag{A.16}$$

where $\|.\|$ is 2-norm defined for a matrix Q as $\|Q\| \triangleq \max_{\|x\|=1} \|Qx\|$ and for a vector $x = [x_1, \ldots, x_n]^T$ as $\|x\| = \sqrt{\sum_{i=1}^n |x_i|^2}$, the matrix $\Omega(s,t)$ can be expanded into a convergent Magnus series $\Omega(t) = \sum_{k=1}^{\infty} \Omega_k(t)$. The convergent condition (A.16) is only sufficient but not necessary. The first term of the

Magnus expansion of $\Omega(s,t)$ is just the exponent of the solution of (A.14) by assuming the commutative property of $Q(t)$, i.e.

$$\Omega_1(s,t) = \int_s^t Q(u)du. \tag{A.17}$$

The second term of the Magnus series is

$$\Omega_2(s,t) = \frac{1}{2}\int_s^t du_1 \int_s^{u_1} du_2[Q(u_1), Q(u_2)]$$

where $[A,B] \triangleq AB - BA$ is the matrix commutator, and the third term is

$$\begin{aligned}\Omega_3(s,t) =& \frac{1}{6}\int_s^t du_1 \int_s^{u_1} du_2 \int_s^{u_2} du_3 \left([Q(u_1)\right. \\ & \left., [Q(u_2), Q(u_3)]] + [Q(u_3), [Q(u_2), Q(u_1)]]\right).\end{aligned}$$

The calculation of further terms can be found in the review [147], which is not involved in this paper.

The transition matrix $P(s,t)$ is time-symmetric in the sense that $P(s,t)P(t,s) = I$, and then $\Omega(s,t) = -\Omega(t,s)$. In the time interval $[s,t]$,

$$\Omega\left(\frac{s+t}{2} - h, \frac{s+t}{2} + h\right) = -\Omega\left(\frac{s+t}{2} + h, \frac{s+t}{2} - h\right)$$

for $h \leq (t-s)/2$ and thus the odd function $\Omega_k\left(\frac{s+t}{2} - h, \frac{s+t}{2} + h\right)$ only contains odd powers of h in its Taylor expansion [147, p. 165]. Thus, $\Omega_{2k}(s,t) = O((t-s)^{2k+1})$ and $\Omega_{2k+1}(s,t) = O((t-s)^{2k+3})$ in the time interval $[s,t]$. Correspondingly, $\sum_{k=1}^{2i+1}\Omega_k(s,t) = \Omega(s,t) + O((t-s)^{2i+3})$ and $\sum_{k=1}^{2i}\Omega_k(s,t) = \Omega(s,t) + O((t-s)^{2i+3})$. The sums of the first $2i$ and $2i+1$ terms of the Magnus series of $\Omega(s,t)$ achieve a same order of accuracy with respect to the time length $t-s$. Moreover, the power series of the matrix exponential indicates that $\exp(\Omega(s,t)+O((t-s)^k)) = \exp(\Omega(s,t))+O((t-s)^k)$. Specifically, we have $P(s,t) = \exp(\Omega_1(s,t)) + O((t-s)^3)$ by only keeping the first term in the Magnus expansion. Using the Taylor expansion of $\Omega_1(s,t)$, we may find that Stroock's method only achieves a second-order accuracy by only preserving the first term of the Taylor series of $Q(t)$.

In NIMFA, given the infection probabilities $v_i(t)$ for $i = 1,\ldots,N$, the infinitesimal generator $Q_j(t)$ of a node j defined by (2.6) is a determined function of time. Thus, we can apply the Magnus expansion to the Markov process (2.5), and assume the transition matrix $P_j(s,t) = \exp(\Omega(s,t;j))$. First,

we derive the length of the convergent time interval of the Magnus expansion by condition (A.16). We may verify that the 2-norm of the matrix $Q_j(t)$ is $\|Q_j(t)\| = \sqrt{\tilde{\beta}_j^2(t) + \delta^2}$. For $t > 0$, $\tilde{\beta}_j(t) = \beta \sum_{i=1}^{N} a_{ji} v_i(t) < \beta d_j$, where d_j is the degree of node j, and consequently $\|Q_j(t)\| < \sqrt{\beta^2 d_j^2 + \delta^2}$. Thus,

$$\int_s^{s+T} \|Q_j(t)\| dt < \sqrt{\beta^2 d_j^2 + \delta^2} T$$

Let $\sqrt{\beta^2 d_j^2 + \delta^2} T = \pi$, and thus

$$T = \frac{\pi}{\sqrt{\beta^2 d_j^2 + \delta^2}}. \tag{A.18}$$

The Magnus expansion of $\Omega(s, t; j)$ always converges if $t - s \leq T$ from Eq. (A.16).

A.3 Appendix of Chapter 4

A.3.1 When the Weibull shape parameter tends to zero

We consider the Weibullian SIS process with an infection rate $\beta > 0$. When $\alpha \to 0$, the distribution function of the infection time T is,

$$\begin{aligned} \lim_{\alpha \to 0} F_T(x) &= \lim_{\alpha \to 0} \left(1 - \mathrm{e}^{-[\beta x \Gamma(1+1/\alpha)]^\alpha}\right) \\ &= 1 - \mathrm{e}^{-\lim\limits_{\alpha \to 0} \Gamma(1+1/\alpha)^\alpha} \end{aligned} \tag{A.19}$$

Since $\Gamma(1+1/\alpha)^\alpha = \mathrm{e}^{\alpha \ln \Gamma(1+1/\alpha)}$, we invoke [206, 6.1.40] the asymptotic formula $\ln \Gamma(z) \sim \left(z - \frac{1}{2}\right) \ln z + O(z)$, and then we obtain

$$\begin{aligned} \alpha \ln \Gamma\left(1 + \frac{1}{\alpha}\right) &= \alpha \left(\frac{1}{\alpha} + \frac{1}{2}\right) \ln\left(\frac{1}{a} + 1\right) + O(1) \\ &= -\ln \alpha - \frac{\alpha}{2} \ln \alpha + O(1) \end{aligned}$$

Thus,

$$\lim_{\alpha \to 0} \Gamma\left(1 + \frac{1}{\alpha}\right)^\alpha = \exp\left[\lim_{\alpha \to 0} \left(-\ln \alpha - \frac{\alpha}{2} \ln \alpha\right)\right] = \infty$$

From the calculation above, $\lim_{\alpha\to 0} F_T(x) = 1$ for $x > 0$, i.e., $\lim_{\alpha\to 0} \Pr[T = 0] = 1$. Thus, when $\alpha \to 0$ and $\beta > 0$, an infected node asymptotically almost surely infects its neighbour consistently and all node will be infected. By a similar method, we can verify that the mean-field threshold $\frac{1}{\Gamma(1+1/\alpha)[\Gamma(\alpha+1)]^{1/\alpha}\lambda_1^{1/\alpha}}$ given in [163] also tends to zero for an arbitrary small α.

A.3.2 Proof of Theorem 4.2.1

Proof. We denote Eq. (4.9) by a function $\Phi : [0,1]^N \to [0,1]^N$ that $\mathbf{v}(n/\beta) = \Phi(\mathbf{v}((n-1)/\beta))$ and $v_j(n/\beta) = \Phi_j(\mathbf{v}((n-1)/\beta))$. We may verify that

$$\left.\frac{\partial \Phi_j(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{0}} = \begin{cases} \mathrm{e}^{-1/\tau} & \text{if } a_{ji} = 1 \text{ or } j = i \\ 0 & \text{if } a_{ji} = 0 \end{cases}$$

which is the element of the Jacobian matrix $J_\Phi(\mathbf{0})$ of the function Φ at $\mathbf{0}$ in the j-th row and i-th column. Thus, the Jacobian matrix is $J_\Phi(\mathbf{0}) = \mathrm{e}^{-1/\tau}(A+I)$, and we assume that λ_{max} is the largest eigenvalue of the Jacobian $J_\Phi(\mathbf{0})$ in absolute value.

Since the network is connected, the matrix $J_\Phi(\mathbf{0})$ is irreducible. Thus, λ_{max} is the largest eigenvalue of $J_\Phi(\mathbf{0})$ by Perron-Frobenius theorem [90], and then

$$\lambda_{\text{max}} = \mathrm{e}^{-1/\tau}(\lambda_1 + 1)$$

For the dynamical system $\mathbf{x}(n) = \Phi(\mathbf{x}(n-1))$ in form (4.9), Ahn and Hassibi [168, Theorem 5.1] have indicated that $\mathbf{0}$ is globally stable and that $\lim_{n\to\infty} \mathbf{x}(n) = \mathbf{0}$ for any $\mathbf{x}(0) \in [0,1]^N$ when $\lambda_{\text{max}} < 1$. While if $\lambda_{\text{max}} > 1$, then there exits one and only one non-zero globally stable point such that $\mathbf{0} \prec \lim_{n\to\infty} \mathbf{x}(n)$ for any $\mathbf{x}(0) \in [0,1]^N$ and $\mathbf{x}(0) \neq \mathbf{0}$. Thus, the maximum infection probability $\mathbf{v}(n/\beta)$ of each time interval governed by Eq. (4.9) converges to $\mathbf{0}$ when $\lambda_{\text{max}} < 1$, and $\mathbf{v}(n/\beta)$ converges to the unique non-zero constant infection probability $\mathbf{v}_\infty(0)$ when $\lambda_{\text{max}} > 1$ and $\mathbf{v}(0) \neq \mathbf{0}$. Thus, $\lambda_{\text{max}} = 1$ is the critical point at which the phase transition happens. Let $\lambda_{\text{max}} = 1$, and we obtain

$$\tau = \frac{1}{\ln(\lambda_1 + 1)}$$

which is the epidemic threshold of the Weibullian SIS process with $\alpha \to \infty$. In each time interval in the steady state, the infection probability is $\mathbf{v}_\infty(t^*) = \mathbf{v}_\infty(0)\mathrm{e}^{-\delta t^*}$, which follows from (4.8). ■

A.3.3 Below the epidemic threshold in the limiting case

From (4.9), the infection probability at $t^* = 0$ of n-th time interval is upper bounded by

$$v_j\left(\frac{n}{\beta}\right) \leq v_j\left(\frac{n-1}{\beta}\right)\mathrm{e}^{-1/\tau} + \left[\sum_{i=1}^{N} a_{ji}v_i\left(\frac{n-1}{\beta}\right)\mathrm{e}^{-1/\tau}\right] \quad \text{(A.20)}$$

From the inequality above, we have

$$\begin{aligned}\mathbf{v}\left(\frac{n}{\beta}\right) &\preceq \mathrm{e}^{-1/\tau}(A+I)\mathbf{v}\left(\frac{n-1}{\beta}\right)\\ &\preceq \left[\mathrm{e}^{-1/\tau}(A+I)\right]^n\mathbf{v}\left(0\right)\\ &= \sum_{i=1}^{N}\left[\mathrm{e}^{-1/\tau}(\lambda_i+1)\right]^n\mathbf{u}_i\mathbf{u}_i^T\mathbf{v}(0)\end{aligned}$$

where $\lambda_1 \geq \cdots \geq \lambda_i \geq \cdots \geq \lambda_N$ are the eigenvalues of the matrix A, and $\mathbf{u}_i$ is the corresponding eigenvector of λ_i. Thus, there exists a constant vector $\mathbf{z}$ where every element is positive, such that,

$$\mathbf{v}\left(\frac{n}{\beta}\right) \preceq \left[\mathrm{e}^{-1/\tau}(\lambda_1+1)\right]^n\mathbf{z} \quad \text{(A.21)}$$

We consider the inequality above for general $t = n/\beta + t^*$. Since $\mathbf{v}(t^* + n/\beta) = \mathbf{v}(n/\beta)\mathrm{e}^{-\delta t^*}$ and $n = \beta(t - t^*)$, we have

$$\begin{aligned}\mathbf{v}(t) = \mathbf{v}\left(\frac{n}{\beta}\right)\mathrm{e}^{-\delta t^*} &\preceq \left[\mathrm{e}^{-1/\tau}(\lambda_1+1)\right]^{\beta(t-t^*)}\mathrm{e}^{-\delta t^*}\mathbf{z}\\ &= \left[\mathrm{e}^{-\delta}(\lambda_1+1)^\beta\right]^{t-t^*}\left(\mathrm{e}^{-\delta}\right)^{t^*}\mathbf{z}\\ &\prec \left[\mathrm{e}^{-\delta}(\lambda_1+1)^\beta\right]^{t}\mathbf{z}\end{aligned}$$

If the effective infection rate τ is below the mean-field epidemic threshold $\tau_c^{(1)}$, then $\mathrm{e}^{-\delta}(\lambda_1+1)^\beta < 1$, and $\left[\mathrm{e}^{-\delta}(\lambda_1+1)^\beta\right]^t\mathbf{z}$ is exponentially decreasing with time t.

A.3.4 Probability generating function of the Weibull distribution

The probability generating function (pgf) $\varphi_T(z) = E\left[e^{-zT}\right]$ of a continuous random variable T is defined as the double sided Laplace transform [88, p. 20]

of the probability density function $f_T(x)$. Clearly, $\varphi_T(0) = 1$. We explore properties of the pgf of the Weibull distribution with pdf,

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

with mean $E[T] = b\Gamma\left(1+\frac{1}{\alpha}\right)$ and distribution function $F_T(t) = \Pr[T \le t]$,

$$F_T(t) = \int_0^t \frac{\alpha}{b}\left(\frac{x}{b}\right)^{\alpha-1} e^{-\left(\frac{x}{b}\right)^{\alpha}} dx = 1 - e^{-\left(\frac{t}{b}\right)^{\alpha}}$$

and pgf

$$\varphi_T(z) = \frac{\alpha}{b}\int_0^{\infty} e^{-zu}\left(\frac{u}{b}\right)^{\alpha-1} e^{-\left(\frac{u}{b}\right)^{\alpha}} du$$

Let $x = \frac{u}{b}$, then we obtain with $w = bz$ and explicitly expressing the dependence on the "shape" parameter α

$$\varphi_T(w;\alpha) = \alpha\int_0^{\infty} e^{-wx-x^{\alpha}} x^{\alpha-1} dx \tag{A.22}$$

which illustrates that the pgf of a Weibull distribution consists of two parameters, $w = zb$ (which is a complex number) and the real non-negative number α. Further, differentiating (A.22)

$$\frac{d\varphi_T(w;\alpha)}{dw} = -\alpha\int_0^{\infty} e^{-wx-x^{\alpha}} x^{\alpha} dx < 0$$

demonstrates, since the integrand is always non-negative, that $\varphi_T(w;\alpha)$ monotonously decreases with w along the real w axis from $\varphi_T(0;\alpha) = 1$ towards $\lim_{w\to\infty}\varphi_T(w;\alpha) = 0$.

After substituting $u = x^{\alpha}$, we find for all $\alpha > 0$ that

$$\varphi_T(w;\alpha) = \int_0^{\infty} e^{-wu^{\frac{1}{\alpha}}} e^{-u} du \tag{A.23}$$

Partial integration of the integral in (A.22) yields

$$\varphi_T(w;\alpha) = 1 - w\int_0^{\infty} e^{-wx} e^{-x^{\alpha}} dx \tag{A.24}$$

Let $u = wx$, then $\varphi_T(w;\alpha) = 1 - \int_0^{\infty} e^{-w^{-\alpha}u^{\alpha}} e^{-u} du$ and comparison with (A.23) leads to a *functional equation* for the pgf of a Weibull random variable T,

$$\varphi_T(w;\alpha) = 1 - \varphi_T\left(\frac{1}{w^{\alpha}};\frac{1}{\alpha}\right) \tag{A.25}$$

Only for a few values of α, the pgf $\varphi_T(w;\alpha)$ can be analytically evaluated. For $\alpha = 1$, the Weibull distribution reduces to an exponential and (A.22) is $\varphi_T(w;1) = \frac{1}{1+w}$. For $\alpha = 2$, we find that $\varphi_T(w;2) = 1 - we^{\frac{w^2}{4}} \int_{\frac{w}{2}}^{\infty} e^{-u^2} du$ and the functional equation (A.25) gives us $\varphi_T\left(w;\frac{1}{2}\right) = 1 - \varphi_T\left(w^{-\frac{1}{2}};2\right) = \frac{e^{\frac{1}{4w}}}{\sqrt{w}} \int_{\frac{1}{2\sqrt{w}}}^{\infty} e^{-u^2} du$. From (A.23), we find that $\lim_{\alpha\to\infty} \varphi_T(w;\alpha) = e^{-w}$ and $\lim_{\alpha\to 0} \varphi_T(w;\alpha) = 0$ for $\mathrm{Re}(w) > 0$, but $\lim_{\alpha\to 0} \varphi_T(w;\alpha) = \infty$ if $\mathrm{Re}(w) < 0$.

A.4 Appendix of Chapter 5

A.4.1 The coefficient of the maximum prevalence

If the adjacency matrix of the network is A, the largest eigenvalue of A is λ_1, the normalized principal eigenvalue of A is $x = [x_1, \ldots, x_N]^T$, and the effective infection rate is $\tau = \beta/\delta$ with infection rate β and curing rate δ, then the epidemic threshold [156, Theorem 1] of the bursty SIS model is $\tau_c^{(B)} = \frac{1}{\ln(\lambda_1+1)}$ and the following Theorem holds.

Theorem A.4.1 For the bursty SIS process with effective infection rate τ above the threshold $\tilde{\tau} \triangleq \frac{\tau}{\tau_c^{(B)}} - 1 > 0$, the maximum steady-state prevalence is $y_\infty^+(\tilde{\tau}) = a_{\max}\tilde{\tau} + o(\tilde{\tau})$ with

$$a_{\max} = \frac{2}{N} \frac{(\lambda_1+1)\ln(\lambda_1+1)\sum_{i=1}^N x_i}{\lambda_1 \sum_{i=1}^N x_i^3}$$

and the minimum prevalence is $y_\infty^-(\tilde{\tau}) = a_{\min}\tilde{\tau} + o(\tilde{\tau})$ with $a_{\min} = a_{\max}/(\lambda_1+1)$.

To prove Theorem A.4.1, we first prove the following Lemma.

Lemma A.4.1

$$\sum_{i=1}^N x_i \sum_{\{j,k\in\mathcal{N}_i | j<k\}} x_j x_k + \lambda_1 \sum_{i=1}^N x_i^3 = \frac{1}{2}\lambda_1(\lambda_1+1)\sum_{i=1}^N x_i^3$$

where $\mathcal{N}_i$ denotes the set of neighbors of node i.

Proof of Lemma A.4.1. For the first term on the left-hand side, we have

$$\begin{aligned}\sum_{i=1}^{N} x_i \sum_{\{j,k\in\mathcal{N}_i | j<k\}} x_j x_k &= \frac{1}{2}\sum_{i=1}^{N} x_i \sum_{j\in\mathcal{N}_i} x_j \left(\sum_{k\in\mathcal{N}_i} x_k - x_j\right) \\ &= \frac{1}{2}\sum_{i=1}^{N} x_i \sum_{j\in\mathcal{N}_i} x_j \sum_{k\in\mathcal{N}_i} x_k \\ &\quad -\frac{1}{2}\sum_{i=1}^{N} x_i \sum_{j\in\mathcal{N}_i} x_j^2 \end{aligned} \tag{A.26}$$

Since $\sum_{j\in\mathcal{N}_i} x_j = \lambda_1 x_i$, the first term of (A.26) is $\frac{1}{2}\lambda_1^2 \sum_{i=1}^{N} x_i^3$. We consider the second term of (A.26)

$$\begin{aligned}-\frac{1}{2}\sum_{i=1}^{N} x_i \sum_{j\in\mathcal{N}_i} x_j^2 &= -\frac{1}{2}\sum_{\forall \text{link}(i,j)} \left(x_i^2 x_j + x_i x_j^2\right) \\ &= -\frac{1}{2}\sum_{i=1}^{N} x_i^2 \sum_{j\in\mathcal{N}_i} x_j \\ &= -\frac{1}{2}\lambda_1 \sum_{i=1}^{N} x_i^3 \end{aligned}$$

Thus, the left-hand side equals $\frac{1}{2}\lambda_1(\lambda_1+1)\sum_{i=1}^{N} x_i^3$. ■

Proof of Theorem A.4.1. The mean-field governing equations of the bursty SIS process are [156],

$$\begin{aligned}v_i\left(\frac{n+1}{\beta}\right) = \lim_{t^*\to 1/\beta} \Bigg(&\left[1 - v_i\left(t^* + \frac{n}{\beta}\right)\right]\left(1 - \prod_{j\in\mathcal{N}_i}\left[1 - v_j\left(t^* + \frac{n}{\beta}\right)\right]\right) \\ &+ v_i\left(t^* + \frac{n}{\beta}\right)\Bigg)\end{aligned} \tag{A.27}$$

and

$$\frac{\mathrm{d}v_i\left(\frac{n}{\beta} + t^*\right)}{\mathrm{d}t^*} = -\delta v_i\left(\frac{n}{\beta} + t^*\right) \tag{A.28}$$

where $v_i(t)$ is the infection probability of node i at time t, $t^* \in [0, 1/\beta)$ is the length of the time passed after the nearest burst, and $\mathcal{N}_i$ denotes the set of

neighbor nodes of node i. The solution of Eq. (A.28) is

$$v_i\left(\frac{n}{\beta}+t^*\right)=v_i\left(\frac{n}{\beta}\right)\mathrm{e}^{-\delta t^*} \tag{A.29}$$

Substituting (A.29) at $t^* \rightarrow 1/\beta$, i.e. $\lim_{t^*\rightarrow 1/\beta} v_i(n/\beta+t^*) = v_i(n/\beta)\exp(-1/\tau)$, into Eq. (A.27), we obtain the following recursion of the infection probability of each node at $t^*=0$ just after each burst,

$$\begin{aligned} v_i\left(\frac{n+1}{\beta}\right)=&\left[1-v_i\left(\frac{n}{\beta}\right)\mathrm{e}^{-1/\tau}\right]\left(1-\prod_{j\in\mathcal{N}_i}\left[1-v_i\left(\frac{n}{\beta}\right)\mathrm{e}^{-1/\tau}\right]\right)\\ &+v_j\left(\frac{n}{\beta}\right)\mathrm{e}^{-1/\tau} \end{aligned} \tag{A.30}$$

Equation (A.30) is the discrete-time SIS equation with infection probability $\tilde{\beta}=\mathrm{e}^{-1/\tau}$ and curing probability $\tilde{\delta}=1-\mathrm{e}^{-1/\tau}$. We rewrite Eq. (A.30) as,

$$p_i[n+1]=\Big(1-(1-\tilde{\delta})p_i[n]\Big)\left[1-\prod_{j\in\mathcal{N}_i}\Big(1-\tilde{\beta}p_j[n]\Big)\right]+p_j[n](1-\tilde{\delta})$$

where $p_i[n]\triangleq v_i(n/\beta)$. In the steady state, $\lim_{n\rightarrow\infty} p_i[n+1]=\lim_{n\rightarrow\infty} p_i[n]=p_{i\infty}$ for $1\leq i\leq N$, and we have,

$$\tilde{\delta}p_{i\infty}=\Big[1-(1-\tilde{\delta})p_{i\infty}\Big]\left[1-\prod_{j\in\mathcal{N}_i}\Big(1-\tilde{\beta}p_{j\infty}\Big)\right] \tag{A.31}$$

In the steady state, the discrete-time SIS infection probability vector $p_\infty\triangleq[p_{1\infty},\ldots,p_{N\infty}]$ approaches an eigenvector of the adjacency matrix A corresponding to the largest eigenvalue λ_1 when $\tilde{\beta}/\tilde{\delta}\downarrow 1/\lambda_1$. Thus, we can assume $p_\infty=ax+o(a)q$, where q is a vector orthogonal to x and with finite components.

Substituting $p_\infty=ax+o(a)q$ into (A.31), we obtain,

$$\begin{aligned} \tilde{\delta}ax_i+\tilde{\delta}o(a)q_i=&\tilde{\beta}a\sum_{j\in\mathcal{N}_i}x_j+\tilde{\beta}o(a)\sum_{j\in\mathcal{N}_j}q_j-a^2\tilde{\beta}^2\sum_{\{j,k\in\mathcal{N}_i|j<k\}}x_jx_k\\ &-\tilde{\beta}(1-\tilde{\delta})a^2x_i\sum_{j\in\mathcal{N}_i}x_j+o(a^2) \end{aligned} \tag{A.32}$$

where the eigenvalue equation indicates that $\sum_{j\in\mathcal{N}_i} x_j = \lambda_1 x_i$.

In vector form, (A.32) is,

$$\begin{aligned}\tilde{\delta}ax + \tilde{\delta}o(a)q =&\tilde{\beta}aAx + \tilde{\beta}o(a)Aq - a^2\tilde{\beta}^2\text{vec}\left(\sum_{\{j,k\in\mathcal{N}_i|j<k\}} x_j x_k\right)\\ &- \tilde{\beta}(1-\tilde{\delta})a^2\text{vec}\left(\lambda_1 x_i^2\right) + o(a^2)h\end{aligned} \tag{A.33}$$

where the vector $\text{vec}(z_i) \triangleq [z_1, \ldots, z_N]^T$. Divide both sides of (A.33) by $a\tilde{\beta}$ and recall that $Ax = \lambda_1 x$, and we have

$$\begin{aligned}\frac{\tilde{\delta}}{\tilde{\beta}}x + \frac{\tilde{\delta}}{\tilde{\beta}}\frac{o(a)}{a}q =&\lambda_1 x + \frac{o(a)}{a}Aq - a\tilde{\beta}\text{vec}\left(\sum_{\{j,k\in\mathcal{N}_i|j<k\}} x_j x_k\right)\\ &- a(1-\tilde{\delta})\text{vec}\left(\lambda_1 x_i^2\right) + \frac{o(a^2)}{a}h\end{aligned} \tag{A.34}$$

Rearranging (A.34), we obtain

$$\begin{aligned}\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)x - \frac{\tilde{\delta}}{\tilde{\beta}}\frac{o(a)}{a}q =& -\frac{o(a)}{a}Aq + a\Bigg[\tilde{\beta}\text{vec}\left(\sum_{\{j,k\in\mathcal{N}_i|j<k\}} x_j x_k\right)\\ &+ (1-\tilde{\delta})\text{vec}\left(\lambda_1 x_i^2\right)\Bigg] + \frac{o(a^2)}{a}h\end{aligned} \tag{A.35}$$

Since $a \to 0$ as $(\lambda_1 - \tilde{\delta}/\tilde{\beta}) \to 0$, we assume

$$a = a_1(\lambda_1 - \tilde{\delta}/\tilde{\beta}) + o(\lambda_1 - \tilde{\delta}/\tilde{\beta}) \tag{A.36}$$

and substitute a into (A.35),

$$\begin{aligned}\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)x - \frac{\tilde{\delta}}{\tilde{\beta}}\frac{o(a)}{a}q =& -\frac{o(a)}{a}Aq + a_1\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)d(\tilde{\beta},\tilde{\delta})\\ &+ o(a)d(\tilde{\beta},\tilde{\delta}) + \frac{o(a^2)}{a}h\end{aligned} \tag{A.37}$$

where

$$d(\tilde{\beta},\tilde{\delta}) = \left(\tilde{\beta}\text{vec}\left(\sum_{\{j,k\in\mathcal{N}_i|j<k\}} x_j x_k\right) + (1-\tilde{\delta})\text{vec}\left(\lambda_1 x_i^2\right)\right).$$

We divide both side of Eq. (A.37) by $\lambda_1 - \tilde{\delta}/\tilde{\beta}$,

$$\begin{aligned} x - \frac{\tilde{\delta}}{\tilde{\beta}}\frac{o(a)}{a}\frac{1}{\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)}q = &- \frac{o(a)}{a}\frac{1}{\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)}Aq + a_1 d(\tilde{\beta}, \tilde{\delta}) \\ &+ \frac{o(a)}{\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)}d(\tilde{\beta}, \tilde{\delta}) + \frac{o(a^2)}{a\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)}h \end{aligned} \tag{A.38}$$

By taking the scalar product with x on both sides of Eq. (A.38) and recalling that the vector q is orthogonal to the eigenvector x, we obtain

$$1 = a_1 d(\tilde{\beta}, \tilde{\delta}) \cdot x + \frac{o(a)}{\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)}d(\tilde{\beta}, \tilde{\delta}) \cdot x + \frac{o(a^2)}{a\left(\lambda_1 - \frac{\tilde{\delta}}{\tilde{\beta}}\right)}h \cdot x \tag{A.39}$$

When $a \to 0$, Eq. (A.39) becomes

$$1 = a_1 d\left(\frac{1}{\lambda_1}\tilde{\delta}, \tilde{\delta}\right) \cdot x \tag{A.40}$$

In the bursty SIS case where $\lim_{\tau \downarrow \tau_c^{(B)}} \tilde{\delta} = \frac{\lambda_1}{\lambda_1 + 1}$, Eq. (A.40) reads

$$1 = a_1 d\left(\frac{1}{\lambda_1 + 1}, \frac{\lambda_1}{\lambda_1 + 1}\right) \cdot x$$

Thus,

$$a_1 = \frac{\lambda_1 + 1}{\sum_{i=1}^{N} x_i \sum_{\{j,k \in \mathcal{N}_i | j<k\}} x_j x_k + \lambda_1 \sum_{i=1}^{N} x_i^3}$$

Using Lemma A.4.1, a_1 becomes

$$a_1 = \frac{2}{\lambda_1 \sum_{i=1}^{N} x_i^3}$$

We assume $a = a_2\epsilon + o(\epsilon)$ where $\epsilon = \tau - \tau_c^{(B)} = \tau - \frac{1}{\ln(\lambda_1 + 1)}$ and we may verify that

$$\left.\frac{\mathrm{d}(\lambda_1 - \tilde{\delta}/\tilde{\beta})}{\mathrm{d}\epsilon}\right|_{\epsilon=0} = \left.\frac{\mathrm{d}(\lambda_1 + 1 - \mathrm{e}^{1/\tau})}{\mathrm{d}\epsilon}\right|_{\epsilon=0} = (\lambda_1 + 1)\ln^2(\lambda_1 + 1)$$

then we obtain

$$a_2 = a_1(\lambda_1 + 1)\ln^2(\lambda_1 + 1) = \frac{2(\lambda_1 + 1)\ln^2(\lambda_1 + 1)}{\lambda_1 \sum_{i=1}^{N} x_i^3}$$

Thus, the maximum prevalence is $\frac{a_2 \sum_{i=1}^{N} x_i}{N} \left(\tau - \tau_c^{(B)}\right) + o(\tau - \tau_c^{(B)})$.

After normalizing the effective infection rate by $\tau/\tau_c^{(B)}$ and defining $\tilde{\tau} = \tau/\tau_c^{(B)} - 1$, we finally find the maximum prevalence as

$$\begin{aligned} y_\infty^+(\tilde{\tau}) &= \frac{a_2 \tau_c^{(B)} \sum_{i=1}^{N} x_i}{N} \tilde{\tau} + o(\tilde{\tau}) \\ &= \frac{2(\lambda_1+1)\ln(\lambda_1+1)\sum_{i=1}^{N} x_i}{N\lambda_1 \sum_{i=1}^{N} x_i^3} \tilde{\tau} + o(\tilde{\tau}) \end{aligned} \tag{A.41}$$

For general t^*, the prevalence is $\exp(-\delta t^*) y_\infty^+(\tilde{\tau})$ and then the minimum prevalence is $y_\infty^-(\tilde{\tau}) = y_\infty^+(\tilde{\tau})/(\lambda_1+1)$ as $t^* \to 1/\beta$. ■

A.4.2 The bounds of the coefficient of NIMFA prevalence

By the Perron-Frobenius theorem, every component of the principal eigenvector is positive. The lower bound of a is derived follows.

$$a = \frac{\sum_{i=1}^{N} x_i}{N \sum_{i=1}^{N} x_i^3} \geq \frac{N \min_i x_i}{N \max_i x_i \sum_{j=1}^{N} x_j^2} = \frac{\min_i x_i}{\max_i x_i}$$

For the upper bound, using the Cauchy-Schwarz inequality $(\sum_{i=1}^{N} x_i)^2 \leq N \sum_{i=1}^{N} x_i^2 = N$, we obtain

$$a = \frac{\sum_{i=1}^{N} x_i}{N \sum_{i=1}^{N} x_i^3} \leq \frac{\sqrt{N}}{N \min_i x_i} = \frac{1}{\sqrt{N} \min_i x_i}$$

The bound is tight when the network is a regular graph.

A.4.3 The coefficients of d-regular graphs

For regular graph, the principal eigenvector is $x = \frac{1}{\sqrt{N}} u$ where u is all-one vector and the largest eigenvalue is d. We may verify that

$$\begin{aligned} a &= \frac{\sum_{i=1}^{N} x_i}{N \sum_{i=1}^{N} x_i^3} = 1 \\ a_{\max} &= \frac{2}{N} \frac{(\lambda_1+1)\ln(\lambda_1+1)\sum_{i=1}^{N} x_i}{\lambda_1 \sum_{i=1}^{N} x_i^3} = 2\left(1 + \frac{1}{d}\right)\ln(d+1) \\ a_{\min} &= \frac{2}{N} \frac{\ln(\lambda_1+1)\sum_{i=1}^{N} x_i}{\lambda_1 \sum_{i=1}^{N} x_i^3} = \frac{2\ln(d+1)}{d} \end{aligned}$$

A.4.4 The coefficients of star graphs

We may verify that the largest eigenvalue of the star graph is $\sqrt{N-1}$ and the principle eigenvector is $x = [\frac{1}{\sqrt{2}}, \ldots, \frac{1}{\sqrt{2(N-1)}}]^T$. We have following results

$$
\begin{aligned}
a &= \frac{\sum_{i=1}^{N} x_i}{N \sum_{i=1}^{N} x_i^3} = \frac{1}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right) \\
a_{\max} &= \frac{2}{N} \frac{(\lambda_1 + 1)\ln(\lambda_1 + 1) \sum_{i=1}^{N} x_i}{\lambda_1 \sum_{i=1}^{N} x_i^3} = \frac{\ln(\sqrt{N})}{\sqrt{N}} + o\left(N^{-\frac{1}{2}} \ln N\right) \\
a_{\min} &= \frac{2}{N} \frac{\ln(\lambda_1 + 1) \sum_{i=1}^{N} x_i}{\lambda_1 \sum_{i=1}^{N} x_i^3} = \frac{\ln(\sqrt{N})}{N} + o\left(N^{-1} \ln N\right)
\end{aligned}
$$

A.4.5 Real networks

The parameters of the real and synthetic networks are listed in Table A.1. The degree distributions are plotted in Fig. A.1 to Fig. A.7.

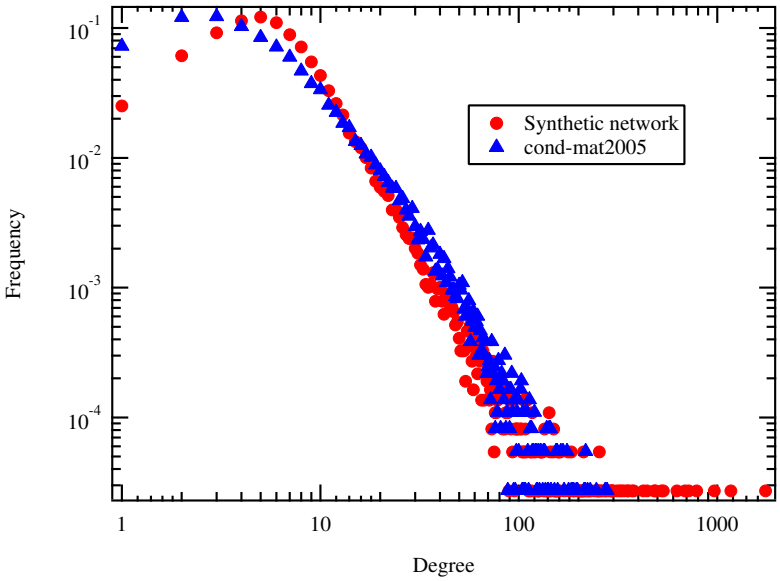

Figure A.1: Cond-mat2005: Collaboration network of scientists posting preprints on the condensed matter archive at arXiv, 1995-1999.

Networks	Real networks	Synthetic networks
Cond-mat 2005 [33]	$N = 36458$ $d_{av} = 9.4210$ $a_{\max} = 0.1199$	$N = 36811$ $d_{av} = 9.4483$ $a_{\max} = 0.1301$
astro-ph [33]	$N = 14845$ $d_{av} = 16.1202$ $a_{\max} = 0.3024$	$N = 14766$ $d_{av} = 16.2536$ $a_{\max} = 0.5352$
Internet [192]	$N = 22963$ $d_{av} = 4.2186$ $a_{\max} = 0.2155$	$N = 22354$ $d_{av} = 4.2804$ $a_{\max} = 0.1903$
hep-th [33]	$N = 5835$ $d_{av} = 4.7352$ $a_{\max} = 0.0218$	$N = 5944$ $d_{av} = 4.5855$ $a_{\max} = 0.1063$
Email-URV [190]	$N = 1133$ $d_{av} = 9.6222$ $a_{\max} = 1.3713$	$N = 1178$ $d_{av} = 9.6774$ $a_{\max} = 0.9539$
PGP [191]	$N = 10680$ $d_{av} = 4.5536$ $a_{\max} = 0.0789$	$N = 10986$ $d_{av} = 4.4773$ $a_{\max} = 0.2104$
Email-Enron [193]	$N = 33696$ $d_{av} = 10.7319$ $a_{\max} = 0.3037$	$N = 33632$ $d_{av} = 10.8451$ $a_{\max} = 0.3053$

Table A.1: The parameters of real networks and the corresponding synthetic networks. Only the largest connected components are preserved and all the networks are connected.

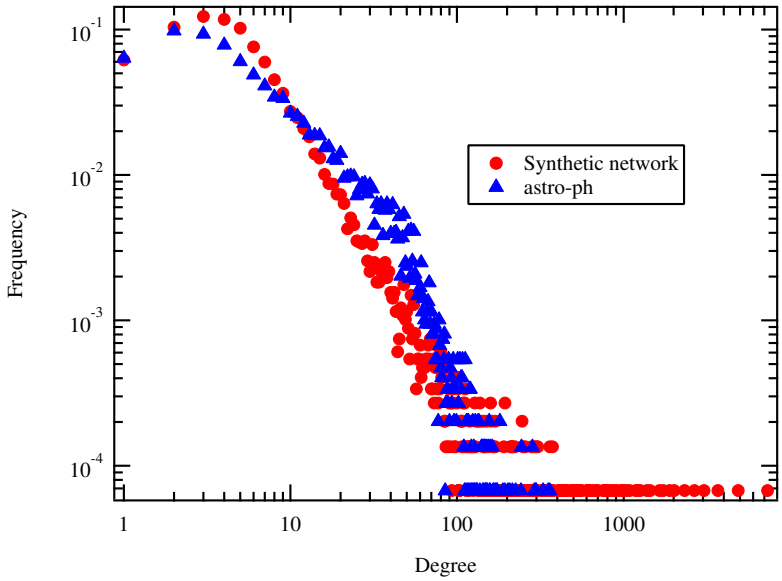

Figure A.2: Astro-ph: Network of co-authorship between scientists posting preprints on the Astrophysics E-Print Archive between Jan 1, 1995 and December 31, 1999.

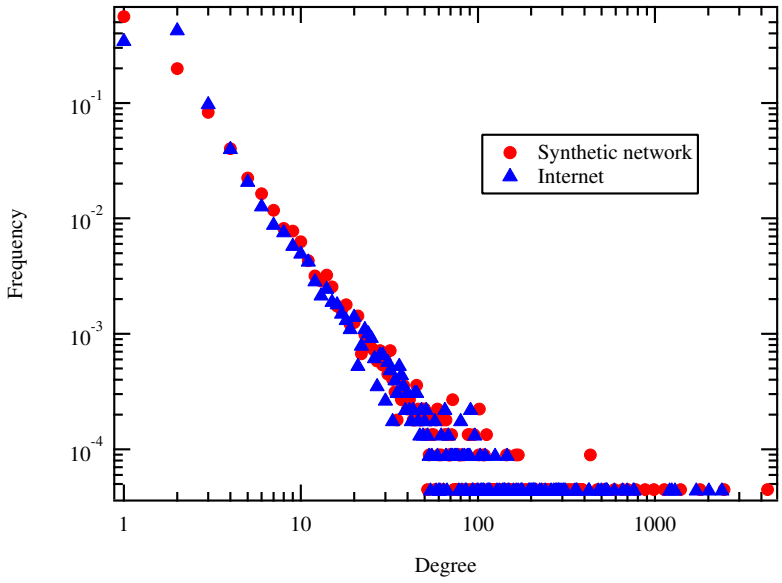

Figure A.3: Internet: a symmetrized snapshot of the structure of the Internet at the level of autonomous systems, reconstructed from BGP tables posted by the University of Oregon Route Views Project.

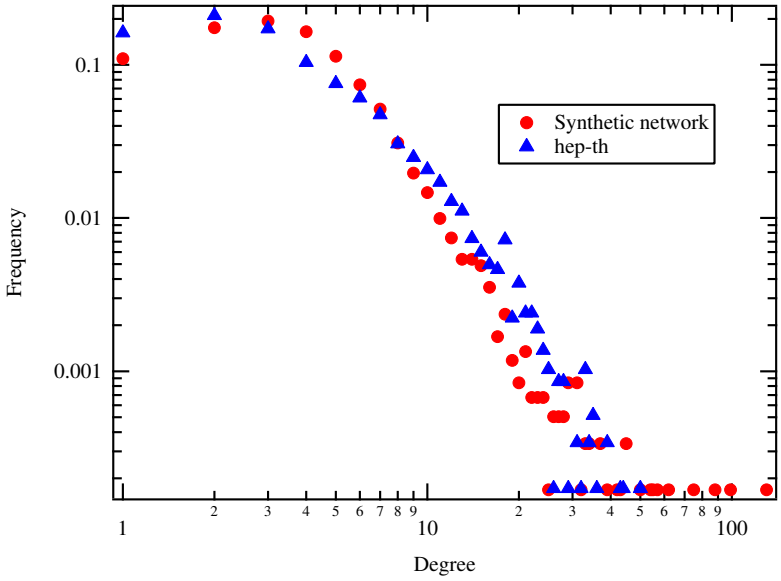

Figure A.4: Hep-th: Network of co-authorship between scientists posting preprints on the High-Energy Theory arXiv between Jan 1, 1995 and December 31, 1999.

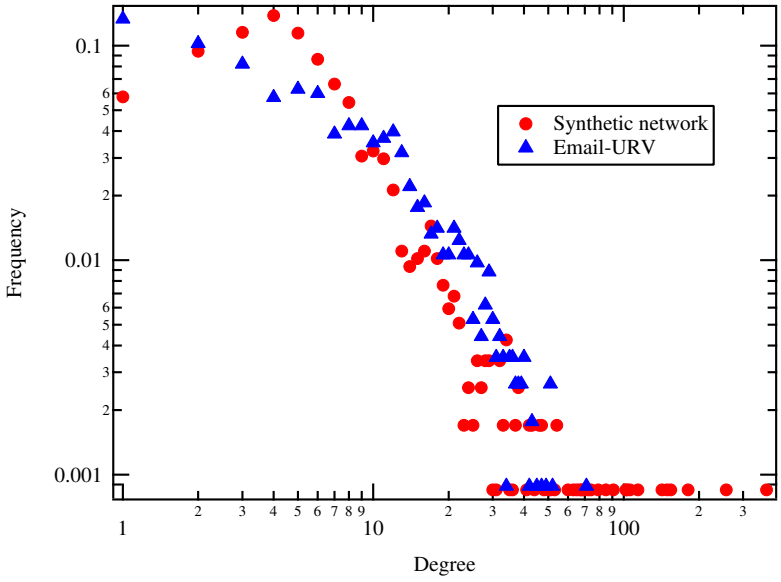

Figure A.5: Email-URV: Network of E-mail interchanges between members of the Univeristy Rovira i Virgili, Tarragona.

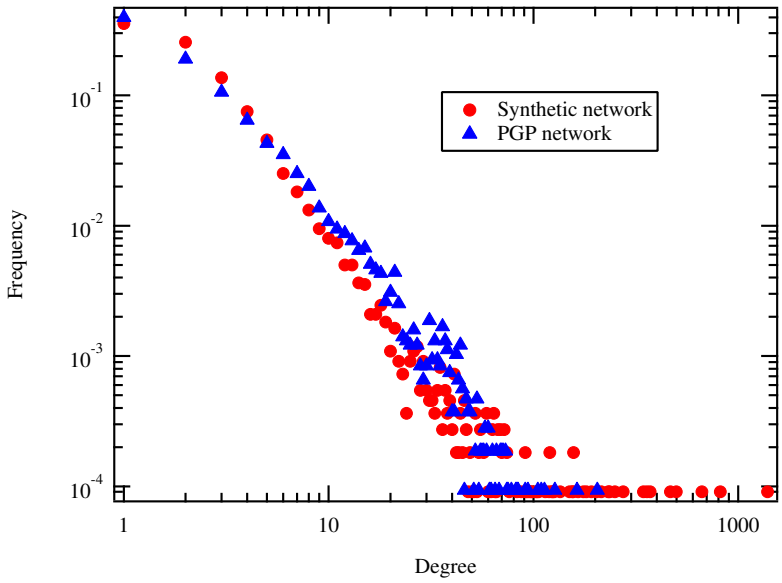

Figure A.6: PGP: Network of users of the Pretty-Good-Privacy algorithm for secure information interchange.

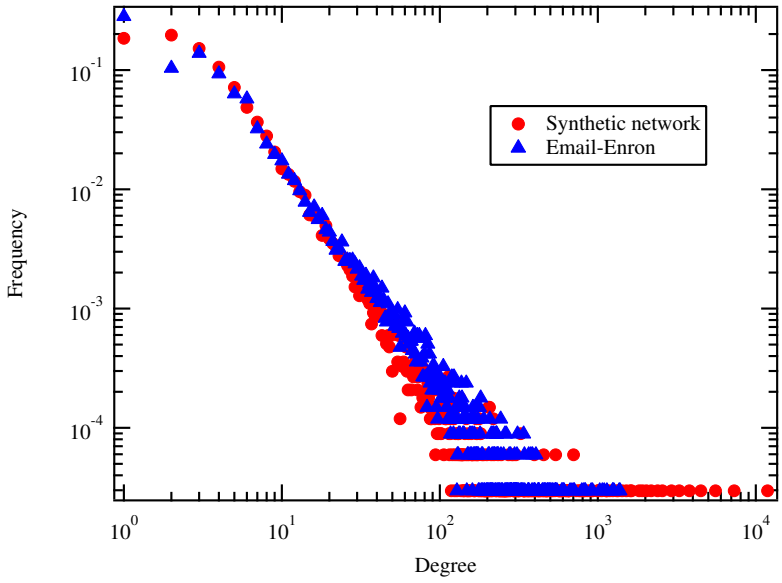

Figure A.7: Email-Enron: Enron email communication network.

Bibliography

[1] P. W. Anderson, *More is different,* Science **177**, 393 (1972).

[2] D. Bernoulli and S. Blower, *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it,* Reviews in Medical Virology **14**, 275 (2004).

[3] D. Bernoulli, *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir,* Mem. Math. Phys. Acad. R. Sci. , 1 (1766).

[4] W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics,* Proceedings of the royal society of London A **115**, 700 (1927).

[5] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control* (Oxford university press, 1992).

[6] T. E. Harris, *Contact interactions on a lattice,* Ann. Probab. **2**, 969 (1974).

[7] T. M. Liggett, *Stochastic interacting systems: contact, voter and exclusion processes*, Vol. 324 (springer science & Business Media, 2013).

[8] J. Snow, *On the mode of communication of cholera* (John Churchill, 1855).

[9] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, and T. Başar, *Analysis, estimation, and validation of discrete-time epidemic processes,* IEEE Transactions on Control Systems Technology , 1 (2018).

[10] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Critical phenomena in complex networks,* Reviews of Modern Physics **80**, 1275 (2008).

[11] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic processes in complex networks,* Reviews of Modern Physics **87**, 925 (2015).

[12] R. Pastor-Satorras and A. Vespignani, *Epidemic spreading in scale-free networks,* Physical Review Letters **86**, 3200 (2001).

[13] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks,* Science **286**, 509 (1999).

[14] H. K. Lee, P.-S. Shim, and J. D. Noh, *Epidemic threshold of the Susceptible-Infected-Susceptible model on complex networks,* Physical Review E **87**, 062812 (2013).

[15] W. Cota, S. C. Ferreira, and G. Ódor, *Griffiths effects of the Susceptible-Infected-Susceptible epidemic model on random power-law networks,* Physical Review E **93**, 032322 (2016).

[16] M. A. Muñoz, R. Juhász, C. Castellano, and G. Ódor, *Griffiths phases on complex networks,* Physical Review Letters **105**, 128701 (2010).

[17] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. Mendes, *Localization and spreading of diseases in complex networks,* Physical Review Letters **109**, 128702 (2012).

[18] R. S. Morison, *Is there a biological person?* The Milbank Memorial Fund Quarterly. Health and Society **61**, 3 (1983).

[19] P. Moretti and M. A. Muñoz, *Griffiths phases and the stretching of criticality in brain networks,* Nature Communications **4**, 2521 (2013).

[20] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, *Critical points and transitions in an electric power transmission model for cascading failure blackouts,* Chaos **12**, 985 (2002).

[21] M. E. Nickol and J. Kindrachuk, *A year of terror and a century of reflection: perspectives on the great influenza pandemic of 1918–1919,* BMC Infectious Diseases **19**, 117 (2019).

[22] P. Bak, C. Tang, and K. Wiesenfeld, *Self-organized criticality: An explanation of the 1/f noise,* Physical Review Letters **59**, 381 (1987).

[23] J. Scargill, *Can life exist in 2+ 1 dimensions?* arXiv: 1906.05336 (2019).

[24] Q. Liu and P. Van Mieghem, *Autocorrelation of the susceptible-infected-susceptible process on networks,* Physical Review E **97**, 062309 (2018).

[25] Q. Liu and P. Van Mieghem, *Evaluation of an analytic, approximate formula for the time-varying SIS prevalence in different networks,* Physica A **471**, 325 (2017).

[26] Q. Liu and P. Van Mieghem, *Die-out probability in sis epidemic processes on networks,* in *Complex Networks & Their Applications V*, edited by H. Cherifi, S. Gaito, W. Quattrociocchi, and A. Sala (Springer International Publishing, Cham, 2017) pp. 511–521.

[27] S. H. Strogatz, *Exploring complex networks,* Nature **410**, 268 (2001).

[28] G. A. Pagani and M. Aiello, *The power grid as a complex network: A survey,* Physica A **392**, 2688 (2013).

[29] H. Cetinay, K. Devriendt, and P. Van Mieghem, *Nodal vulnerability to targeted attacks in power grids,* Applied Network Science **3**, 34 (2018).

[30] H. Cetinay, S. Soltan, F. A. Kuipers, G. Zussman, and P. Van Mieghem, *Comparing the effects of failures in power grids under the ac and dc power flow models,* IEEE Transactions on Network Science and Engineering **5**, 301 (2018).

[31] H. Cetinay, F. A. Kuipers, and P. Van Mieghem, *A topological investigation of power flow,* IEEE Systems Journal **12**, 2524 (2018).

[32] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the internet topology,* SIGCOMM Comput. Commun. Rev. **29**, 251 (1999).

[33] M. E. J. Newman, *The structure of scientific collaboration networks,* Proceedings of the national academy of sciences **98**, 404 (2001).

[34] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, *Economic networks: The new challenges,* Science **325**, 422 (2009).

[35] W. W. Zachary, *An information flow model for conflict and fission in small groups,* Journal of Anthropological Research **33**, 452 (1977).

[36] S. Tang, N. Blenn, C. Doerr, and P. Van Mieghem, *Digging in the digg social news website,* IEEE Transactions on Multimedia **13**, 1163 (2011).

[37] A. R. Mashaghi, A. Ramezanpour, and V. Karimipour, *Investigation of a protein complex network,* The European Physical Journal B **41**, 113 (2004).

[38] E. Bullmore and O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems,* Nature reviews neuroscience **10**, 186 (2009).

[39] S. J. Cook, T. A. Jarrell, C. A. Brittin, Y. Wang, A. E. Bloniarz, M. A. Yakovlev, K. C. Q. Nguyen, L. T.-H. Tang, E. A. Bayer, J. S. Duerr, H. E. Bülow, O. Hobert, D. H. Hall, and S. W. Emmons, *Whole-animal connectomes of both caenorhabditis elegans sexes,* Nature **571**, 63–71 (2019).

[40] C. J. Stam, A. Hillebrand, H. Wang, and P. Van Mieghem, *Emergence of modular structure in a large-scale brain network with interactions between dynamics and connectivity,* Frontiers in Computational Neuroscience **4**, 133 (2010).

[41] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Random graphs with arbitrary degree distributions and their applications,* Phys. Rev. E **64**, 026118 (2001).

[42] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem, *Distances in random graphs with finite variance degrees,* Random Structures & Algorithms **27**, 76 (2005).

[43] G. Hooghiemstra and P. Van Mieghem, *On the mean distance in scale free graphs,* Methodology and Computing in Applied Probability **7**, 285 (2005).

[44] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world'networks,* nature **393**, 440 (1998).

[45] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks,* Proceedings of the National Academy of Sciences **99**, 7821 (2002).

[46] M. E. J. Newman, *Assortative mixing in networks,* Physical Review Letters **89**, 208701 (2002).

[47] D. C. Mocanu, E. Mocanu, P. Stone, P. H. Nguyen, M. Gibescu, and A. Liotta, *Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science,* Nature communications **9**, 2383 (2018).

[48] S. Xie, A. Kirillov, R. Girshick, and K. He, *Exploring randomly wired neural networks for image recognition,* arXiv: 1904.01569 (2019).

[49] X. Zheng, F. C. M. Lau, C. K. Tse, Y. He, and S. Hau, *Application of complex-network theories to the design of short-length low-density-parity-check codes,* IET Communications **3**, 1569 (2009).

[50] X. Zheng, F. C. M. Lau, and C. K. Tse, *Constructing high-rate scale-free ldpc codes,* in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (2010) pp. 3781–3784.

[51] B. Karrer and M. E. J. Newman, *Stochastic blockmodels and community structure in networks,* Physical Review E **83**, 016107 (2011).

[52] T. P. Peixoto, *Nonparametric bayesian inference of the microcanonical stochastic block model,* Physical Review E **95**, 012317 (2017).

[53] M. Newman, *Network structure from rich but noisy data,* Nature Physics **14**, 542 (2018).

[54] T. P. Peixoto, *Reconstructing networks with unknown and heterogeneous errors,* Physical Review X **8**, 041011 (2018).

[55] T. Martin, B. Ball, and M. E. J. Newman, *Structural inference for uncertain networks,* Physical Review E **93**, 012306 (2016).

[56] M. E. J. Newman, *Estimating network structure from unreliable measurements,* Physical Review E **98**, 062321 (2018).

[57] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, *Geometric deep learning: Going beyond euclidean data,* IEEE Signal Processing Magazine **34**, 18 (2017).

[58] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, *The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,* IEEE Signal Processing Magazine **30**, 83 (2013).

[59] L. Euler, *Solutio problematis ad geometriam situs pertinentis,* Commentarii academiae scientiarum Petropolitanae , 128 (1741).

[60] J. H. Barnett, *Early writings on graph theory: Euler circuits and the königsberg bridge problem,* (2005).

[61] P. Erdös and A. Rényi, *On random graphs, i,* Publicationes Mathematicae (Debrecen) **6**, 290 (1959).

[62] J. Park and M. E. J. Newman, *Statistical mechanics of networks,* Physical Review E **70**, 066117 (2004).

[63] A. Fronczak, P. Fronczak, and J. A. Hołyst, *Average path length in random networks,* Physical Review E **70**, 056110 (2004).

[64] F. Karinthy, *Chain-links,* Everything is different (1929).

[65] S. Milgram, *The small world problem,* Psychology today **2**, 60 (1967).

[66] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, *The anatomy of the facebook social graph,* arXiv: 1111.4503 (2011).

[67] M. E. J. Newman, *Scientific collaboration networks. i. network construction and fundamental results,* Physical Review E **64**, 016131 (2001).

[68] M. E. J. Newman, *Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality,* Physical Review E **64**, 016132 (2001).

[69] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, *The web of human sexual contacts,* Nature **411**, 907 (2001).

[70] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks,* Reviews of Modern Physics **74**, 47 (2002).

[71] R. Albert, H. Jeong, and A.-L. Barabási, *Internet: Diameter of the world-wide web,* Nature **401**, 130 (1999).

[72] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, *The large-scale organization of metabolic networks,* Nature **407**, 651 (2000).

[73] M. E. J. Newman, *Power laws, pareto distributions and zipf's law,* Contemporary Physics **46**, 323 (2005).

[74] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (OUP Oxford, 2013).

[75] C. Dangalchev, *Generation models for scale-free networks,* Physica A **338**, 659 (2004).

[76] M. K. Hassan, L. Islam, and S. A. Haque, *Degree distribution, rank-size distribution, and leadership persistence in mediation-driven attachment networks,* Physica A **469**, 23 (2017).

[77] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Connectivity of growing random networks,* Physical Review Letters **85**, 4629 (2000).

[78] E. Ravasz and A.-L. Barabási, *Hierarchical organization in complex networks,* Physical Review E **67**, 026112 (2003).

[79] R. Cohen and S. Havlin, *Scale-free networks are ultrasmall,* Physical Review Letters **90**, 058701 (2003).

[80] A. D. Broido and A. Clauset, *Scale-free networks are rare,* Nature communications **10**, 1017 (2019).

[81] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, *Scale-free networks well done,* arXiv: 1811.02071 (2018).

[82] A. Barabási, *Love is all you need: Clauset's fruitless search for scale-free networks,* Blog post available at https://www. barabasilab. com/post/love-is-all-you-need (2018).

[83] P. Holme, *Rare and everywhere: Perspectives on scale-free networks,* Nature communications **10**, 1016 (2019).

[84] M. Newman, *Networks: an introduction* (Oxford university press, 2010).

[85] A.-L. Barabási *et al.*, *Network science* (Cambridge university press, 2016).

[86] R. van der Hofstad, *Random Graphs and Complex Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 1 (Cambridge University Press, 2016).

[87] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).

[88] P. Van Mieghem, *Performance analysis of complex networks and systems* (Cambridge University Press, Cambridge, 2014).

[89] D. Easley, J. Kleinberg, *et al.*, *Networks, crowds, and markets*, Vol. 8 (Cambridge university press Cambridge, 2010).

[90] P. Van Mieghem, *Graph spectra for complex networks* (Cambridge University Press, 2010).

[91] P. Van Mieghem and Q. Liu, *Explicit non-markovian susceptible-infected-susceptible mean-field epidemic threshold for weibull and gamma infections but poisson curings,* Phys. Rev. E **100**, 022317 (2019).

[92] F. Chung, L. Lu, and V. Vu, *Eigenvalues of random power law graphs,* Annals of Combinatorics **7**, 21 (2003).

[93] P. Van Mieghem, *A new type of lower bound for the largest eigenvalue of a symmetric matrix,* Linear Algebra and its Applications **427**, 119 (2007).

[94] S. Walker and P. V. Mieghem, *On lower bounds for the largest eigenvalue of a symmetric matrix,* Linear Algebra and its Applications **429**, 519 (2008).

[95] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, and H. Wang, *Decreasing the spectral radius of a graph by link removals,* Phys. Rev. E **84**, 016101 (2011).

[96] M. E. J. Newman, *Mathematics of networks,* in *The New Palgrave Dictionary of Economics* (Palgrave Macmillan UK, London, 2016) pp. 1–8.

[97] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web.*, Technical Report 1999-66 (Stanford InfoLab, 1999).

[98] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, *Localization of the maximal entropy random walk,* Physical Review Letters **102**, 160602 (2009).

[99] N. Masuda, M. A. Porter, and R. Lambiotte, *Random walks and diffusion on networks,* Physics Reports **716-717**, 1 (2017).

[100] M. Fiedler, *Matrices and Graphs in Geometry*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, 2011).

[101] K. Devriendt and P. Van Mieghem, *The simplex geometry of graphs,* Journal of Complex Networks (2019), 10.1093/comnet/cny036.

[102] P. Van Mieghem, J. Omic, and R. Kooij, *Virus Spread in Networks,* IEEE/ACM Transactions on Networking **17**, 1 (2009).

[103] P. L. Simon, M. Taylor, and I. Z. Kiss, *Exact epidemic models on graphs using graph-automorphism driven lumping,* Journal of mathematical biology **62**, 479 (2011).

[104] P. Van Mieghem, *Decay towards the overall-healthy state in sis epidemics on networks,* arXiv: 1310.3980 (2013).

[105] P. Van Mieghem, *Time evolution of SIS epidemics in the complete graph,* Delft University of Technology, Report 20170405, https://www.nas.ewi.tudelft.nl/people/Piet/TUDelftReports.html (2017).

[106] P. Van Mieghem, F. D. Sahnehz, and C. Scoglioz, *An upper bound for the epidemic threshold in exact markovian sir and sis epidemics on networks,* in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on* (IEEE, 2014) pp. 6228–6233.

[107] P. Van Mieghem, *The N-intertwined SIS epidemic network model,* Computing **93**, 147 (2011).

[108] C. Castellano and R. Pastor-Satorras, *Thresholds for epidemic spreading in networks,* Physical Review Letters **105**, 218701 (2010).

[109] P. Van Mieghem, *An epidemic perspective on the cut size in networks,* Delft University of Technology, Report 20180312, https://www.nas.ewi.tudelft.nl/people/Piet/TUDelftReports.html (2018).

[110] E. Cator and P. Van Mieghem, *Nodal infection in markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated,* Physical Review E **89**, 052802 (2014).

[111] P. M. Rodríguez, A. Roldán-Correa, and L. A. Valencia, *Comment on "nodal infection in markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated",* Physical Review E **98**, 026301 (2018).

[112] E. Cator, P. Donnelly, and P. Van Mieghem, *Reply to "comment on 'nodal infection in markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated' ",* Physical Review E **98**, 026302 (2018).

[113] P. Donnelly, *The correlation structure of epidemic models,* Mathematical biosciences **117**, 49 (1993).

[114] A. Khanafer, T. Basar, and B. Gharesifard, *Stability properties of infected networks with low curing rates,* in *American Control Conference* (IEEE, 2014) pp. 3579–3584.

[115] P. Van Mieghem and R. van de Bovenkamp, *Non-markovian infection spread dramatically alters the Susceptible-Infected-Susceptible epidemic threshold in networks,* Physical Review Letters **110**, 108701 (2013).

[116] K. T. D. Eames and M. J. Keeling, *Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases,* Proceedings of the National Academy of Sciences **99**, 13330 (2002).

[117] C. Li, R. van de Bovenkamp, and P. Van Mieghem, *Susceptible-infected-susceptible model: A comparison of n-intertwined and heterogeneous mean-field approximations,* Physical Review E **86**, 026116 (2012).

[118] M. Boguñá and R. Pastor-Satorras, *Epidemic spreading in correlated complex networks,* Physical Review E **66**, 047104 (2002).

[119] J. P. Gleeson, *High-accuracy approximation of binary-state dynamics on networks,* Physical Review Letters **107**, 068701 (2011).

[120] J. P. Gleeson, *Binary-state dynamics on complex networks: Pair approximation and beyond,* Physical Review X **3**, 021004 (2013).

[121] A. S. Mata and S. C. Ferreira, *Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks,* EPL (Europhysics Letters) **103**, 48003 (2013).

[122] E. Cator and P. Van Mieghem, *Second-order mean-field susceptible-infected-susceptible epidemic threshold,* Physical Review E **85**, 056111 (2012).

[123] M. E. J. Newman, *Spread of epidemic disease on networks,* Physical Review E **66**, 016128 (2002).

[124] E. Kenah and J. M. Robins, *Second look at the spread of epidemics on networks,* Physical Review E **76**, 036113 (2007).

[125] B. Karrer and M. E. J. Newman, *Message passing approach for general epidemic models,* Physical Review E **82**, 016101 (2010).

[126] R. Parshani, S. Carmi, and S. Havlin, *Epidemic threshold for the susceptible-infectious-susceptible model on random networks,* Physical Review Letters **104**, 258701 (2010).

[127] M. Shrestha, S. V. Scarpino, and C. Moore, *Message-passing approach for recurrent-state epidemic models on networks,* Physical Review E **92**, 022821 (2015).

[128] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, *Epidemic spreading in real networks: an eigenvalue viewpoint,* in *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.* (2003) pp. 25–34.

[129] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, *Epidemic thresholds in real networks,* ACM Transactions on Information and System Security **10**, 1 (2008).

[130] A. Lajmanovich and J. A. Yorke, *A deterministic model for gonorrhea in a nonhomogeneous population,* Mathematical Biosciences **28**, 221 (1976).

[131] D. T. Gillespie, A. Hellander, and L. R. Petzold, *Perspective: Stochastic algorithms for chemical kinetics,* The Journal of Chemical Physics **138**, 170901 (2013).

[132] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions,* The Journal of Physical Chemistry **81**, 2340 (1977).

[133] M. Boguñá, L. F. Lafuerza, R. Toral, and M. A. Serrano, *Simulating non-markovian stochastic processes,* Physical Review E **90**, 042108 (2014).

[134] W. Cota and S. C. Ferreira, *Optimized gillespie algorithms for the simulation of markovian epidemic processes on large and heterogeneous networks,* Computer Physics Communications **219**, 303 (2017).

[135] G. St-Onge, J.-G. Young, L. Hébert-Dufresne, and L. J. Dubé, *Efficient sampling of spreading processes on complex networks using a composition and rejection algorithm,* Computer Physics Communications **240**, 30 (2019).

[136] R. van de Bovenkamp, *Epidemic Processes on Complex Networks: Modelling, Simulation and Algorithms*, Ph.D. thesis, TU Delft, Delft University of Technology (2015).

[137] I. Z. Kiss, J. C. Miller, P. L. Simon, *et al.*, *Mathematics of epidemics on networks,* Cham: Springer (2017).

[138] J. C. Miller, *Mathematics-of-epidemics-on-networks,* `https://github.com/springer-math/Mathematics-of-Epidemics-on-Networks` (2018).

[139] P. Van Mieghem and R. van de Bovenkamp, *Accuracy criterion for the mean-field approximation in susceptible-infected-susceptible epidemics on networks,* Physical Review E **91**, 032812 (2015).

[140] E. Cator and P. Van Mieghem, *Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis,* Physical Review E **87**, 012811 (2013).

[141] R. van de Bovenkamp and P. Van Mieghem, *Survival time of the susceptible-infected-susceptible infection process on a graph,* Physical Review E **92**, 032806 (2015).

[142] P. Van Mieghem, *Approximate formula and bounds for the time-varying Susceptible-Infected-Susceptible prevalence in networks,* Physical Review E **93**, 052312 (2016).

[143] R. M. Anderson, B. T. Grenfell, and R. M. May, *Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis,* Journal of Hygiene **93**, 587–608 (1984).

[144] J. Meier, X. Zhou, A. Hillebrand, P. Tewarie, C. Stam, and P. Van Mieghem, *The epidemic spreading model and the direction of information flow in brain networks,* NeuroImage **152**, 639 (2017).

[145] C. Li, R. van de Bovenkamp, and P. Van Mieghem, *Susceptible-Infected-Susceptible model: A comparison of N-Intertwined and heterogeneous mean-field approximations,* Physical Review E **86**, 026116 (2012).

[146] W. Magnus, *On the exponential solution of differential equations for a linear operator,* Communications on pure and applied mathematics **7**, 649 (1954).

[147] S. Blanes, F. Casas, J. Oteo, and J. Ros, *The Magnus expansion and some of its applications,* Physics Reports **470**, 151 (2009).

[148] W. Cota and S. C. Ferreira, *Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks,* Computer Physics Communications **219**, 303 (2017).

[149] P. Van Mieghem and R. van de Bovenkamp, *Accuracy criterion for the mean-field approximation in Susceptible-Infected-Susceptible epidemics on networks,* Physical Review E **91**, 032812 (2015).

[150] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, *Accuracy of mean-field theory for dynamics on real-world networks,* Physical Review E **85**, 026106 (2012).

[151] K. Devriendt and P. Van Mieghem, *Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality,* Physical Review E **96**, 052314 (2017).

[152] R. Pastor-Satorras and C. Castellano, *Eigenvector localization in real networks and its implications for epidemic spreading,* Journal of Statistical Physics **173**, 1110 (2018).

[153] F. D. Sahneh, A. Vajdi, and C. Scoglio, *Delocalized epidemics on graphs: A maximum entropy approach,* in *American Control Conference (ACC), 2016* (IEEE, 2016) pp. 7346–7351.

[154] R. Ferreira, R. da Costa, S. Dorogovtsev, and J. Mendes, *Metastable localization of diseases in complex networks,* Physical Review E **94**, 062305 (2016).

[155] P. Van Mieghem, *Epidemic phase transition of the SIS type in networks,* EPL (Europhysics Letters) **97**, 48004 (2012).

[156] Q. Liu and P. Van Mieghem, *Burst of virus infection and a possibly largest epidemic threshold of non-markovian susceptible-infected-susceptible processes on networks,* Physical Review E **97**, 022309 (2018).

[157] C. Doerr, N. Blenn, and P. Van Mieghem, *Lognormal infection times of online information spread,* PloS one **8**, e64349 (2013).

[158] Åke Svensson, *A note on generation times in epidemic models,* Mathematical Biosciences **208**, 300 (2007).

[159] J. C. Heijne, P. Teunis, G. Morroy, C. Wijkmans, S. Oostveen, E. Duizer, M. Kretzschmar, and J. Wallinga, *Enhanced hygiene measures and norovirus transmission during an outbreak,* Emerging infectious diseases **15**, 24 (2009).

[160] H. Götz, K. Ekdahl, J. Lindbäck, B. de Jong, K. O. Hedlund, and J. Giesecke, *Clinical Spectrum and Transmission Characteristics of Infection with Norwalk-Like Virus: Findings from a Large Community Outbreak in Sweden,* Clinical Infectious Diseases **33**, 622 (2001).

[161] B. J. Cowling, V. J. Fang, S. Riley, J. M. Peiris, and G. M. Leung, *Estimation of the serial interval of influenza,* Epidemiology **20**, 344 (2009).

[162] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, and M. Murray, *Transmission dynamics and control of severe acute respiratory syndrome,* Science **300**, 1966 (2003).

[163] E. Cator, R. van de Bovenkamp, and P. Van Mieghem, *Susceptible-infected-susceptible epidemics on networks with general infection and cure times,* Physical Review E **87**, 062816 (2013).

[164] A. Vazquez, B. Racz, A. Lukacs, and A.-L. Barabasi, *Impact of non-poissonian activity patterns on spreading processes,* Physical Review Letters **98**, 158702 (2007).

[165] J. W. Tang, K. L. Ngai, W. Y. Lam, and P. K. Chan, *Seasonality of influenza a (h3n2) virus: a hong kong perspective (1997–2006),* PLoS one **3**, e2768 (2008).

[166] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds, *Multiscale, resurgent epidemics in a hierarchical metapopulation model,* Proceedings of the National Academy of Sciences **102**, 11157 (2005).

[167] Z.-D. Zhao, Y. Liu, and M. Tang, *Epidemic variability in hierarchical geographical networks with human activity patterns,* Chaos **22**, 023150 (2012).

[168] H. J. Ahn and B. Hassibi, *Global dynamics of epidemic spread over complex networks,* in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on* (IEEE, 2013) pp. 4579–4585.

[169] C. I. Del Genio, T. Gross, and K. E. Bassler, *All scale-free networks are sparse,* Physical Review Letters **107**, 178701 (2011).

[170] X. R. Hoffmann and M. Boguñá, *Memory-induced complex contagion in epidemic spreading,* New Journal of Physics **21**, 033034 (2019).

[171] Q. Liu and P. Van Mieghem, *Network localization is unalterable by infections in bursts,* IEEE Transactions on Network Science and Engineering (2018), 10.1109/TNSE.2018.2889539.

[172] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, *Gene expression dynamics in the macrophage exhibit criticality,* Proceedings of the National Academy of Sciences **105**, 1897 (2008).

[173] C. Furusawa and K. Kaneko, *Adaptation to optimal cell growth through self-organized criticality,* Physical Review Letters **108**, 208103 (2012).

[174] M. G. Kitzbichler, M. L. Smith, S. R. Christensen, and E. Bullmore, *Broadband criticality of human brain network synchronization,* PLoS computational biology **5**, e1000314 (2009).

[175] M. A. Muñoz, *Colloquium: Criticality and dynamical scaling in living systems,* Reviews of Modern Physics **90**, 031001 (2018).

[176] P. W. Anderson, *Absence of diffusion in certain random lattices,* Physical Review **109**, 1492 (1958).

[177] T. Martin, X. Zhang, and M. E. J. Newman, *Localization and centrality in networks,* Physical Review E **90**, 052808 (2014).

[178] M. Sade, T. Kalisky, S. Havlin, and R. Berkovits, *Localization transition on complex networks via spectral statistics,* Physical Review E **72**, 066123 (2005).

[179] L. Jahnke, J. W. Kantelhardt, R. Berkovits, and S. Havlin, *Wave localization in complex networks with high clustering,* Physical Review Letters **101**, 175702 (2008).

[180] K. J. Sharkey, *Localization of eigenvector centrality in networks with a cut vertex,* Physical Review E **99**, 012315 (2019).

[181] G. F. de Arruda, E. Cozzo, T. P. Peixoto, F. A. Rodrigues, and Y. Moreno, *Disease localization in multilayer networks,* Physical Review X **7**, 011014 (2017).

[182] R. B. Griffiths, *Nonanalytic behavior above the critical point in a random ising ferromagnet,* Physical Review Letters **23**, 17 (1969).

[183] E. C. Titchmarsh, D. R. Heath-Brown, *et al.*, *The theory of the Riemann zeta-function* (Oxford University Press, 1986).

[184] S. Chatterjee, R. Durrett, *et al.*, *Contact processes on random graphs with power law degree distributions have critical value 0,* The Annals of Probability **37**, 2332 (2009).

[185] M. Boguná, C. Castellano, and R. Pastor-Satorras, *Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks,* Physical Review Letters **111**, 068701 (2013).

[186] R. Pastor-Satorras and C. Castellano, *Distinct types of eigenvector localization in networks,* Scientific reports **6**, 18847 (2016).

[187] C. Castellano and R. Pastor-Satorras, *Relating topological determinants of complex networks to their spectral properties: Structural and dynamical effects,* Physical Review X **7**, 041024 (2017).

[188] K.-I. Goh, B. Kahng, and D. Kim, *Universal behavior of load distribution in scale-free networks,* Physical Review Letters **87**, 278701 (2001).

[189] M. P. Stumpf and M. A. Porter, *Critical truths about power laws,* Science **335**, 665 (2012).

[190] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, *Self-similar community structure in a network of human interactions,* Physical Review E **68**, 065103 (2003).

[191] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, *Models of social networks based on social distance attachment,* Physical Review E **70**, 056122 (2004).

[192] M. E. J. Newman, *Mark Newman's network data,* `http://www-personal.umich.edu/~mejn/netdata/` (2013).

[193] J. Leskovec and A. Krevl, *SNAP Datasets: Stanford large network dataset collection,* `http://snap.stanford.edu/data` (2014).

[194] Q. Liu, X. Zhou, and P. V. Mieghem, *Pulse strategy for suppressing spreading on networks,* EPL (Europhysics Letters) **127**, 38001 (2019).

[195] F. S. Dawood, A. D. Iuliano, C. Reed, M. I. Meltzer, D. K. Shay, P.-Y. Cheng, D. Bandaranayake, R. F. Breiman, W. A. Brooks, P. Buchy, D. R. Feikin, K. B. Fowler, A. Gordon, N. T. Hien, P. Horby, Q. S. Huang, M. A. Katz, A. Krishnan, R. Lal, J. M. Montgomery, K. Mølbak, R. Pebody, A. M. Presanis, H. Razuri, A. Steens, Y. O. Tinoco, J. Wallinga, H. Yu, S. Vong, J. Bresee, and M.-A. Widdowson, *Estimated global mortality associated with the first 12 months of 2009 pandemic influenza a h1n1 virus circulation: a modelling study,* The Lancet Infectious Diseases **12**, 687 (2012).

[196] Cisco Systems Inc., *Cisco 2018 annual cybersecurity report,* https://www.cisco.com/c/en/us/products/security/security-reports.html (2018).

[197] S. Vosoughi, D. Roy, and S. Aral, *The spread of true and false news online,* Science **359**, 1146 (2018).

[198] A. Dabbagh, R. L. Laws, C. Steulet, L. Dumolard, M. N. Mulders, K. Kretsinger, J. P. Alexander, P. A. Rota, and J. L. Goodson, *Progress toward regional measles elimination — worldwide, 2000–2017,* MMWR. Morbidity and Mortality Weekly Report **67**, 1323 (2018).

[199] Z. Agur, L. Cojocaru, G. Mazor, R. M. Anderson, and Y. L. Danon, *Pulse mass measles vaccination across age cohorts,* Proceedings of the National Academy of Sciences **90**, 11698 (1993).

[200] L. Stone, B. Shulgin, and Z. Agur, *Theoretical examination of the pulse vaccination policy in the sir epidemic model,* Mathematical and Computer Modelling **31**, 207 (2000), proceedings of the Conference on Dynamical Systems in Biology and Medicine.

[201] Ministry of Health and Family Welfare of India, *Operational guide for pulse polio immunization in india,* http://www.indiaenvironmentportal.org.in/content/389767/operational-guide-for-pulse-polio-immunization-in-india/ (2006).

[202] D. J. Nokes and J. Swinton, *Vaccination in pulses: a strategy for global eradication of measles and polio?* Trends in Microbiology **5**, 14 (1997).

[203] C. R. Doering, K. V. Sargsyan, and L. M. Sander, *Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation,* Multiscale Modeling & Simulation **3**, 283 (2005).

[204] G. St-Onge, J.-G. Young, E. Laurence, C. Murphy, and L. J. Dubé, *Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks,* Physical Review E **97**, 022305 (2018).

[205] D. W. Stroock, *An introduction to Markov processes*, Vol. 230 (Springer Science & Business Media, 2013).

[206] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Vol. 55 (Courier Corporation, 1964).

List of Acronyms and Symbols

SI	Susceptible-Infected
SIS	Susceptible-Infected-Susceptible
NIMFA	The N-Intertwined Mean-Field Approximation
HMF	The Heterogeneous Mean-Field (approximation)
ER	Erdös-Rényi (graphs)
WS	Watts-Strogatz (small-world networks)
BA	Barabási-Albert (scale-free networks)
IPR	Inverse Participant Ratio
A	Adjacency matrix
a_{ij}	The i,j-th element of A
d_i	Degree of node i
d_{av}	Average degree
D	Degree of a randomly selected node
λ_i	The i-th largest eigenvalue of the adjacency matrix
$\mathcal{N}_i$	The set of neighbors of node i
β	Infection rate of each infected node to each healthy neighbor
δ	Curing rate of each infected node
τ	Effective infection rate β/δ
τ_c	Epidemic threshold
$\tau_c^{(1)}$	Mean-field threshold
$\tau_c^{(B)}$	Mean-field threshold of the bursty SIS model
$\tilde{\tau}$	Normalized effective infection rate, e.g. $\tau/\tau_c^{(1)} - 1$ or $\tau/\tau_c^{(B)} - 1$
$X_i(t)$	Infection state of node i at time t, a Bernoulli random variable
$I(t)$	Fraction of infected nodes at time t
$y(t)$	Prevalence, i.e. $E[I(t)]$
$I^{(1)}(t)$	NIMFA fraction of infected nodes at time t
$y^{(1)}(t)$	NIMFA prevalence $E[I^{(1)}(t)]$
$y_\infty(\tilde{\tau})$	NIMFA prevalence in the steady state with $\tilde{\tau}$
$y_\infty^+(\tilde{\tau})$	Maximum mean-field prevalence in the steady state of the bursty SIS model
$y_\infty^-(\tilde{\tau})$	Minimum mean-field prevalence in the steady state of the bursty SIS model

$v_i(t)$	Mean-field infection probability of node i at time t
$v_{i\infty}$	Mean-field infection probability of node i in the steady state
$V_i(t)$	NIMFA infection state of node i at time t
$\Gamma(x)$	Gamma function
$\zeta(x)$	Riemann zeta function

Acknowledgments

My promotor Prof. Piet Van Mieghem always states that the four-year Ph.D. study is very short. Now I agree because, even after four years, the first impression of the Netherlands still appears vividly in my head. When I stepped out of Schiphol airport, the clouds were thick and low. It seemed that everyone could easily reach them by straightening arms and go into them by standing on tiptoe. Standing on the street, I feel a bit nervous about the distinct culture of this unfamiliar tiny country along with the forthcoming Ph.D. life. During the four years, I have learned to write a dissertation for the doctorate and have become used to multicultural environments. I cannot find the same feeling of nervousness about research or freshness about the country anymore, which I view as a sign of successful graduation of the life here, even though it was a nice feeling that I miss. Now, after too much beer has been absorbed, my Ph.D. study at TU Delft comes to an end, but the completion of this dissertation consists of many direct and indirect contributions from other people, to whom I would like to express my sincere gratitude.

I would like to thank my promotor Prof. Piet Van Mieghem for admitting me as his Ph.D. student and guiding me through my research. Without his supervision, I cannot finish this dissertation and would never know how to do research. Although professor is a busy job in a modern university, Piet is still with great enthusiasm when diving into the details of my research, reacts fast and provides valuable feedback to every report. Thanks also to my master supervisor Prof. Zhiping Shi for encouraging me to do a Ph.D. study. I am also grateful to colleagues in the research community who have inspired me or provided insights during discussions through Email or in conferences, including Dr. Joel Miller, Prof. Sergey Dorogovtsev, Prof. Li Chen, Prof. Haoxiang Xia, Yanjie Xu, Qing Yao, and Nanxin Wei. Thanks to Ruud van de Bovenkamp for the simulation tools from which I have also obtained a lot of insights. Thanks also to Dr. Edgar van Boven, Dr. Marc de Crom, Bastian Prasse, and Gabriel Budel for their great efforts in creating the *Samenvatting* of this dissertation. It is also my honor to have Prof. Caterina Scoglio, Prof. Yamir Moreno, Prof. Marián Boguñá, Prof. Rob Kooji, Prof. Alan Hanjalic, and Dr. Johan Dubbeldam as my committee members.

Thanks to other TU Delft colleagues who have made a great research atmosphere during my stay in the Network Architectures and Services group and shared lunch and borrel time, including Dr. Fernando Kuipers, Dr. Hui-

juan Wang, Dr. Edgar van Boven, Dr. Remco Litjens, Dr. Eric Smeitink, Dr. Norbert Blenn, Dr. Farabi Iqbal, Dr. Xiangrong Wang, Dr. Niels van Adrichem, Dr. Marcus Märtens, Dr. Hale Çetinay, Dr. Chuan Lin, Dr. Jaron Sanders, Dr. Bo Qu, Walter Knoop, Joyce van Velzen, Misa Taguchi, Zhidong He, Xiuxiu Zhan, Dr. Cunquan Qu, Peng Sun, Rogier Noldus, Ivan Jokić, Fenghua Wang, Gabriel Budel, Yue Tang, Pascal Lagerweij, and Kumar Navneet, especially my office-mates Dr. Jil Meier, Aleksandar Jacimovic, Negar Ahmadi, Dr. Luxing Yang, Karel Devriendt, Bastian Prasse, Long Ma, Maria Raftopoulou, and Albert Senén Cerdà, my master students Yingli Ni, Aziz Hamad, and Xiaoyu Zhou, colleagues in other groups Dr. Wenjie Pei, Dr. Yazhou Yang, Lingling Lao, Arwa Alsadi, Jintao Yu, Yancong Lin, Gerd Kiene, Troya Köylü, and Abid Moueddene. Many of you are not only colleagues who helped my research, but also friends enriching my daily life out of the university.

Many thanks to my friends who have made my life different and supported me during my Ph.D. study. I would like to thank Xin Du, Rendong Liu, Simin Chen, Jiexiong Tang, and Jian Fang, whom I got to know because of a great plan on a coding project ending up in drinking and eating hotpot every weekend. I would also like to thank Dr. Xiang Fu, Dr. Yue Zhao, Sihang Qiu, Hai Zhu, and Shengzhi Xu for the yearly travelling to the middle of nowhere. Thanks to Xin Jin and Siyun Li who have provided valuable advice during my hard days. Thanks to Mengmeng Ye, Longjian Piao, Meng Meng, Nianlei Zhang, Liangyi Li, Weichen Xu, Joan Hu, Yingfu Xu, and Hanwei Wang from the bouldering group for various exciting activities. Thanks to Huiwen Ding, Linyu Li, and Nandy Bernardino for the fun skating time in the Westblaak park of Rotterdam. Thanks to my roommates Chao Zhang, Xueliang Li, Meng Li, Yifeng Zhao, and Chengkai Dai for their help during my stay, especially to Wei Yuan who took care of me when I was sick. Special thanks to Yanfei Rao (and her cat Mimi) and Peichan Li who have encouraged me a lot during my Ph.D. study. I would also like to thank other friends who form a great part of my life in Delft, including Yilin Wang, Mei Liu, Qian Liu, Jiani Liu, Riming Wang, Hao Yu, Ding Ding, Dr. Xue Yu, Dr. Mengshi Yang, Xinmin You, Tiantian Du, Dr. Yanchen Sun, Dr. Anxiao Zhang, Jin Chang, Feiyu Wang, Renfei Bu, Jie Huang, Qian He, Zheng Chen, Yanruo Sun, Xiaoyu Xie, Sijia Wang, Ladislas van Rijen, He Cheng, Xuehan Jiang, Jian Zheng, Mengyu Zhang, Xun Gong, Hua Fan, Dr. Jianping Wang, Bing Huang, Juan Yan, Dr. Qingqing Ye, Tianchi Tang, Bo Wang, Shiwei Bao, Liangfu Wei, Dr. Jiapeng Yin, Dr. Zhilin Zhang, Dr. Guanliang Chen, Mingzhao Zhuo, Haoxuan Ruan, Qiang Li, Maolong lü, Tongjing Wang, Dr. Xinyuan Mao, Jie Zhang, and Dr.

Zhanxiong Liu. I cannot list them all here and I would like to say sorry to the people that should have been mentioned.

In the end, I would like to thank my family, especially my parents Enduo Liu and Zhaowen Wei, for their support and the freedom that they provide for me to explore my own life. If I have contributed anything to scientific research, the contribution should be attributed to them.

Qiang Liu

October, 2019

Delft, the Netherlands

Curriculum Vitæ

Qiang LIU obtained B.E. in Communication Engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China in 2012. During the undergraduate period, he has participated in many electronic system design projects with peers. After graduation, he joined the National Key Laboratory of Science and Technology on Communications of the same university as a master student. Fascinated by the theories of Claude Shannon and others, he has been involved in many projects on designing and implementing error correction codes in communication systems, including Turbo codes, LDPC codes and Polar codes. He then obtained M.Sc. in Cryptography from UESTC in 2015. In the same year, he started his Ph.D. study at the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, under the supervision of Prof.dr.ir. Piet Van Mieghem, where he started to investigate phenomena introduced by network structures. Apart from his professional life, he has been a skateboarder for approximate ten years.

Publications

Publications

10. P. Van Mieghem and **Q. Liu**, *Explicit non-Markovian susceptible-infected-susceptible mean-field epidemic threshold for Weibull and Gamma infections but Poisson curings*, Physical Review E **100**, 022317, Aug. 2019.

9. **Q. Liu**, X. Zhou and P. Van Mieghem, *Pulse strategy for suppressing spreading on networks*, EPL (Europhysics Letters), **127**, 38001, Sept. 2019.

8. Y. Xu, Y. Si, J. Takekawa, **Q. Liu**, H. H. T. Prins, S. Yin, D. Prosser, P. Gong and W. F. de Boer, *A network approach to prioritize conservation efforts for migratory birds*, Conservation Biology, early access, July 2019.

7. L. Ma, **Q. Liu** and P. Van Mieghem, *Inferring the properties of underlying contact network based on the epidemic prevalence*, Applied Network Science, accepted, Oct. 2019.

6. **Q. Liu** and P. Van Mieghem, *Network localization is unalterable by infections in bursts*, IEEE Transactions on Network Science and Engineering, early access, Dec. 2018.

5. **Q. Liu** and P. Van Mieghem, *Autocorrelation of the susceptible-infected-susceptible process on networks*, Physical Review E **97**, 062309, June 2018.

4. **Q. Liu** and P. Van Mieghem, *Burst of virus infection and a possibly largest epidemic threshold of non-Markovian susceptible-infected-susceptible processes on networks*, Physical Review E **97**, 022309, Feb. 2018.

3. **Q. Liu** and P. Van Mieghem, *Evaluation of an analytic, approximate formula for the time-varying SIS prevalence in different networks*, Physica A **471**, 325-336, Apr. 2017.

2. **Q. Liu** and P. Van Mieghem, *Die-out Probability in SIS Epidemic Processes on Networks*, Fifth International Workshop on Complex Networks and their Applications, Milan, Italy, Nov 30 - Dec 2, 2016.

1. J. Guo, Z. Shi, Z. Liu, Z. Zhang and **Q. Liu**, *Multi-CRC Polar Codes and Their Applications*, IEEE Communications Letters **20** (2), 212-215, Feb. 2016.

Talks

5. Complex Networks: Theory, Methods, and Applications, Lake Como School of Advanced Studies, Como, Italy, May 13-17, 2019.

4. NetSciX18, Hangzhou, China, Jan 5-8, 2018.

3. 6th International Conference on Complex Networks and Their Applications, Lyon, France, Nov 29 - Dec 1, 2017.

2. 5th International Workshop on Complex Networks and Their Applications, Milan, Italy, Nov 30 - Dec 2, 2016.

1. Conference on Complex Systems, Amsterdam, the Netherlands, Sep. 19-22, 2016.


TUDelft