

## Dynamic processes on complex networks

### The role of heterogeneity

Qu, Bo

#### DOI

[10.4233/uuid:b0cec363-1527-4e9a-98de-e85bb94389d8](https://doi.org/10.4233/uuid:b0cec363-1527-4e9a-98de-e85bb94389d8)

#### Publication date

2017

#### Document Version

Final published version

#### Citation (APA)

Qu, B. (2017). *Dynamic processes on complex networks: The role of heterogeneity*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:b0cec363-1527-4e9a-98de-e85bb94389d8>

#### Important note

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

#### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

#### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# **DYNAMIC PROCESSES ON COMPLEX NETWORKS**

THE ROLE OF HETEROGENEITY



# **DYNAMIC PROCESSES ON COMPLEX NETWORKS**

THE ROLE OF HETEROGENEITY

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op dinsdag 5 september 2017 om 10:00 uur

door

**Bo Qu**

Master of Engineering in Computer Systems Organization,  
Shanghai Jiaotong University, Shanghai, China,  
geboren te Harbin, China.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. A. Hanjalic

copromotor: Dr. ir. H. Wang

Samenstelling promotiecommissie:

Rector Magnificus,  
Prof. dr. A. Hanjalic,  
Dr. ir. H. Wang,

voorzitter  
Technische Universiteit Delft  
Technische Universiteit Delft

Onafhankelijke leden:

Prof. dr. C. Scoglio,  
Prof. dr. X. Li,  
Prof. mr. dr. P.H.M. Vervest,  
Prof. dr. ir. D.H.J. Epema,  
Dr. W.M. Ruzsel,  
Prof. dr. ir. M. J. T. Reinders,

Kansas State University  
Fudan University  
Erasmus Universiteit Rotterdam  
Technische Universiteit Delft  
Technische Universiteit Delft  
Technische Universiteit Delft, reservelid



*Keywords:* Complex Networks, Opinion Dynamics, Epidemic Spreading, Directed Networks, Heterogeneous Infection/Recovery rates

*Printed by:* Proefschriftmaken

*Front & Back:* The network in the front cover, containing **35** nodes (of which **19** nodes are filled) and **87** links. The real-world directed network in the back cover shows the author's flight trajectories between Sep. 2016 and Sep. 2017. The "cancellation" on the tulip stamp is designed according to the cancellation Type 30 for pre-paidmail in 1844 in the Netherlands. The date, 14th August, used in the cancellation was the first time the author traveled in Delft as well as the Netherlands in 2011.  
*Special thanks to Wei's help for implementing the design.*

Copyright © 2017 by B. Qu

ISBN 978-94-6186-837-4

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

*To whom I love*



# CONTENTS

<b>Summary</b>	<b>xi</b>
<b>Samenvatting</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.2 Mission and Scope of this Thesis . . . . .	4
1.3 Outline of this Thesis . . . . .	5
1.4 List of publications related to the thesis. . . . .	7
<b>2 Network structures and Dynamic processes</b>	<b>9</b>
2.1 Network Structures . . . . .	10
2.1.1 Erdős-Rényi Random Graphs . . . . .	10
2.1.2 Barabási-Albert Random Graphs. . . . .	10
2.1.3 Configuration Model. . . . .	10
2.2 Dynamic Processes . . . . .	11
2.2.1 Opinion Dynamics. . . . .	11
2.2.2 Epidemic Spreading . . . . .	11
<b>3 Opinion dynamics on directed networks</b>	<b>15</b>
3.1 Introduction . . . . .	16
3.2 Basic definitions and notations . . . . .	17
3.2.1 The directionality $\xi$ and indegree outdegree correlation $\rho$ . . . . .	17
3.3 Algorithm Description . . . . .	17
3.3.1 Directionality-increasing rewiring (DIR) . . . . .	17
3.3.2 Constructing an asymmetric indegree and outdegree network and rewiring it to decrease its directionality (ANC-DDR) . . . . .	18
3.4 The influence of the directionality . . . . .	23
3.5 The influence of indegree and outdegree asymmetry . . . . .	26
3.6 Conclusions. . . . .	29
<b>4 Epidemic Spreading with Heterogeneous Recovery Rates</b>	<b>31</b>
4.1 Introduction . . . . .	32
4.2 Directed networks and the epidemic threshold . . . . .	33
4.2.1 Directed networks . . . . .	33
4.2.2 The epidemic threshold . . . . .	33
4.3 The effect of the heterogeneous recovery rates allocation . . . . .	34
4.3.1 The distribution of the recovery rate $\delta_i$ . . . . .	34
4.3.2 The infection fraction $y_\infty$ vs. $\Re(\lambda_1(-diag(\delta_i) + \beta * A))$ . . . . .	35
4.3.3 The influence of $\alpha$ or $(\alpha_{in}, \alpha_{out})$ . . . . .	37

4.4	Optimal heterogeneous recovery allocations . . . . .	37
4.4.1	The optimal exponents $\alpha$ or $(\alpha_{in}, \alpha_{out})$ . . . . .	37
4.4.2	Optimal heterogeneous vs. homogeneous recovery rates . . . . .	38
4.5	Conclusion . . . . .	40
<b>5</b>	<b>Epidemic Spreading with i.i.d. Heterogeneous Infection Rates</b>	<b>43</b>
5.1	Introduction . . . . .	44
5.2	SIS model with heterogeneous infection rates. . . . .	45
5.2.1	Network construction . . . . .	45
5.2.2	Independent and identically distributed heterogeneous infection rates . . . . .	45
5.2.3	The simulations . . . . .	46
5.3	Small recovery rates. . . . .	47
5.3.1	The observations. . . . .	47
5.3.2	The influence of the moments of the infection rates . . . . .	49
5.3.3	The log-normal distribution vs. the gamma distribution. . . . .	51
5.4	Large recovery rates. . . . .	53
5.5	Real-world networks . . . . .	57
5.5.1	Small recovery rates . . . . .	58
5.5.2	Large recovery rates . . . . .	59
5.6	Discussions . . . . .	61
<b>6</b>	<b>Epidemic Spreading with Correlated Heterogeneous Infection Rates</b>	<b>63</b>
6.1	Introduction . . . . .	64
6.2	Preliminary . . . . .	65
6.2.1	The infection rates . . . . .	65
6.2.2	The simulations . . . . .	66
6.3	Effect on the average fraction $y_\infty$ of infected nodes . . . . .	66
6.3.1	Realistic cases: $\alpha \in [-1, 1]$ . . . . .	66
6.3.2	Extreme cases . . . . .	71
6.4	The wheel network . . . . .	73
6.5	Real-world networks . . . . .	75
6.6	Conclusion . . . . .	77
<b>7</b>	<b>The Accuracy of Mean-Field Approximation</b>	<b>79</b>
7.1	Introduction . . . . .	80
7.2	Preliminary . . . . .	81
7.2.1	The extension of NIMFA . . . . .	81
7.2.2	The i.i.d. heterogeneous infection rates . . . . .	81
7.2.3	The correlated heterogeneous infection rates and the range of $\alpha$ . . . . .	82
7.2.4	The simulations . . . . .	83
7.3	Effect of the heterogeneous infection rates . . . . .	83
7.3.1	The i.i.d. infection rates . . . . .	83
7.3.2	The correlated infection rate . . . . .	85
7.4	Real-world network . . . . .	87
7.5	Conclusion . . . . .	88

<b>8 The Nodal Ranking of Infection Probability</b>	<b>91</b>
8.1 Introduction . . . . .	92
8.2 Results . . . . .	93
8.2.1 The counting of the nodal raking changes . . . . .	93
8.2.2 The total number of crossings in different topologies . . . . .	95
8.2.3 The number of crossings in different intervals of $\tau$ . . . . .	96
8.3 Discussion . . . . .	103
<b>9 Reflections and Recommendations</b>	<b>107</b>
9.1 Main Contributions . . . . .	107
9.2 Future Work . . . . .	109
<b>Bibliography</b>	<b>111</b>
<b>Acknowledgements</b>	<b>121</b>
<b>A The Nodal Ranking of Infection Probability</b>	<b>123</b>
A.1 The crossing behavior of the trajectories $\nu_{k\infty}$ . . . . .	123
A.2 Discussion about the one-crossing assumption . . . . .	123
A.3 The derivation of the lower bound $\chi_l$ . . . . .	125
A.4 Derivatives of $\nu_{i\infty}$ with respect to $\tau$ . . . . .	127
A.5 The value of $\tau_u$ . . . . .	131
A.6 Real-world graphs . . . . .	132
A.7 The comparison between NIFMA and the continuous-time simulation . . . . .	134
A.8 $\sigma^*$ as a function of $\tau$ . . . . .	135
A.9 The Spearman rank correlation $\rho$ as a function of $\kappa$ . . . . .	136
<b>Curriculum Vitæ</b>	<b>137</b>
<b>List of Publications</b>	<b>139</b>



# SUMMARY

In the recent decades, various dynamic process models on complex networks have been built to study the mechanisms by which an opinion, a disease or generally the information spreads in real-world networks. For example, opinion models are developed to illustrate the competition of opinions in a population, and epidemic models are used to describe, e.g. how an epidemic spreads in a social contact network or how information propagates in an online social network. Classic models always assume the homogeneous interactions. For example, the infection rates are the same for all pairs of nodes. However, the infection rates between different pairs of nodes which may depend on e.g. interaction frequencies are usually different, thus heterogeneous. In this thesis, we aim to explore the influence of heterogeneity on dynamic processes especially on the prevalence of an epidemic or opinion. We consider two types of dynamic processes: the Non-Consensus Opinion (NCO) model and the Susceptible-Infected-Susceptible (SIS) model. This thesis is mainly devoted to the latter one. We investigate the heterogeneity in both network topology models, e.g. directed networks, and dynamic process models, such as heterogeneous infection rates.

In Chapter 3, we explore how the heterogeneity in network topology (particularly directed networks) affects the NCO model, especially the critical threshold, i.e., the minimal initial fraction of population for a given opinion such that this opinion survives (forms a giant cluster) in the steady state, where the fraction of an opinion remains stable. We propose two approaches to construct directed networks with different proportion of unidirectional links and different correlation between the indegree and outdegree. We find that networks with more (less) unidirectional (bidirectional) links and a higher indegree and outdegree correlation tend to have a higher critical threshold. Our conclusions indicate the critical role of the directionality and the asymmetry between indegree and outdegree in real-world opinion competitions.

We continue to investigate the SIS epidemic model on directed networks in Chapter 4. We aim to understand how to allocate the limited recovery rates heterogeneously to the nodes such that the prevalence of the epidemic can be reduced. We propose a strategy that assigns each node a recovery rate, which is dependent on the in- and outdegree of that node. In general, our strategy is evidently better than the classic homogeneous allocation of recovery resources in reducing the overall infection, especially when the given recovery resources are sufficient. Our degree based heterogeneous recovery rates allocation strategy illustrates the potential to more effectively reduce infection than the classic homogeneous allocation.

We consider further the heterogeneous infection rates in Chapter 5 and Chapter 6, motivated by real-world datasets. Employing the classic SIS model as the benchmark in Chapter 5, we study the influence of the independently identically distributed infection rates on the average fraction of infected nodes in the metastable state where the infection fraction is nonzero and stable. We find that, when the prevalence is high, the hetero-

geneity of infection rates on average retards the virus spreading and a larger even-order moment of the infection rates leads to a smaller average fraction of infected nodes, but the odd-order moments contribute in the opposite way. However, when the prevalence is low, i.e., the epidemic may die out or infect a small fraction of the population, the heterogeneity may enhance the probability that the epidemic spreads out. Chapter 6 goes one step further than Chapter 5: the heterogeneous infection rate of each pair of nodes is not i.i.d. any more, but correlated with the degrees of its two end nodes. We discover that, when the prevalence of the epidemic is high, a negative correlation between the end node degrees and the infection rate tends to help the epidemic spreading. However, when the prevalence is low, a positive correlation is more likely to enhance the spreading. Our results in Chapter 5 and Chapter 6 shed light on that how the epidemic spreads in the real-world could be far away from classic homogeneous models and reveal the essential role of the heterogeneity in real-world dynamic processes.

In this thesis, the continuous-time simulation is the main approach to study the SIS model with heterogeneous infection or recovery rates. We are interested in how accurate a mean-field approximation could be in such cases. Hence, in Chapter 7, we explore how the heterogeneous infection rates affect the accuracy of NIMFA – an advanced mean-field approximation of SIS model.

The previous chapters as well as most studies on SIS model consider mainly the average fraction of infected nodes, which ignores the infection probability of each node. In Chapter 8, we explore the heterogeneous performance of the nodes, i.e., the infection probability or vulnerability of each node, motivated by the fact that the ranking of the nodal vulnerability given the effective infection rate of an SIS epidemic can be crucial for a network operator to understand which nodes are more vulnerable or should be protected. Via both theoretical and numerical approaches, we unveil that the ranking of nodal vulnerability tends to change more significantly as the effective infection rate varies when the effective infection rate is smaller or in Barabási-Albert than Erdős-Rényi random graphs.

As an initial start, this thesis tries to depict an overview how the prevalence of opinion and epidemic could be influenced by various types of heterogeneity, such as the heterogeneity in dynamic processes (heterogeneous infection or recovery rates) and in network topology. This thesis also inspires more future works. Though various types of heterogeneity are considered in this thesis, the heterogeneity in real-world could be vastly complicated and unexploited. For example, the study of the influence of the temporal heterogeneity, i.e., the infection or recovery rates change and adapt over time, on the prevalence is yet open. Furthermore, how can we design the optimal immunization strategy when different types of heterogeneity are considered?

# SAMENVATTING

In de afgelopen decennia zijn verscheidene modellen van dynamische processen op complexe netwerken opgesteld om de mechanismen te bestuderen hoe een mening, een ziekte of informatie in het algemeen zich verspreidt in de praktijk voorkomende netwerken. Zo zijn opiniemodellen ontwikkeld om de strijd tussen meningen in een populatie te illustreren en worden epidemische modellen gebruikt om bijvoorbeeld te beschrijven hoe een epidemie zich door een sociaal contactennetwerk verspreidt of hoe informatie zich voortplant in een online sociaal netwerk. Klassieke modellen veronderstellen altijd dat interacties homogeen zijn. Een voorbeeld van een klassiek model zou kunnen aannemen dat de kans op infectie gelijk is voor alle knopenparen in een netwerk. De kans op infectie tussen verschillende knopenparen kan echter afhangen van hoe vaak er interacties tussen deze knopen plaatsvinden. Doorgaans is deze interactiefrequentie verschillend per knopenpaar en is er dus sprake van heterogene interacties.

In dit proefschrift onderzoeken we de invloed van heterogeniteit op dynamische processen, in het bijzonder de prevalentie van een epidemie of opinie. Hierin beschouwen we twee soorten dynamische processen: het non-consensusopiniemodel (NCO) en het susceptibel-geïnfecteerd-susceptibelmodel (SIS). Dit proefschrift richt zich voornamelijk op het laatstgenoemde. We onderzoeken de heterogeniteit in zowel netwerktopologiemodellen, bijvoorbeeld gerichte netwerken, als in dynamische procesmodellen, zoals heterogene infectiesnelheden.

In hoofdstuk 3 bekijken we hoe de heterogeniteit in netwerktopologie (voornamelijk gerichte netwerken) invloed heeft op het NCO-model, met name de kritieke drempel (het minimaal benodigde percentage van een populatie dat een gegeven mening aanvankelijk moet delen zodanig dat deze opinie het overleeft, d.w.z. een aanzienlijke cluster vormt, in de stabiele toestand van het netwerk, waarin het percentage van de bevolking dat de mening deelt onveranderd blijft). We stellen twee manieren voor het construeren van gerichte netwerken met verschillende verhoudingen van enkelgerichte verbindingen en verschillende correlaties tussen de ingraad en uitgraad voor. Uit onze bevindingen lijken netwerken met meer enkelgerichte (minder bidirectionele) verbindingen en een grotere correlatie tussen de ingraad en uitgraad een hogere kritieke drempel te hebben. Onze conclusies wijzen op de cruciale rol van de gerichtheid en de asymmetrie tussen de ingraad en uitgraad bij een daadwerkelijke opiniestrijd.

We vervolgen ons onderzoek met het bestuderen van het epidemische SIS-model op gerichte netwerken in hoofdstuk 4. Dit hoofdstuk richt zich op het ontwikkelen van een inzicht hoe de beperkte herstelsnelheden dienen te worden verdeeld over de knooppunten in het netwerk zodanig dat de prevalentie van de epidemie kan worden gereduceerd. We stellen een strategie voor waarbij elke knoop in het netwerk een herstelsnelheid krijgt toegewezen die afhangt van de in- en uitgraad van dat knooppunt. In het algemeen is onze voorgestelde strategie evident beter dan de klassieke, homogene toewijzing van herstelmiddelen in het reduceren van de totale infectiegraad, vooral wanneer de gege-

ven herstellmiddelen toereikend zijn. Onze heterogene strategie voor het toewijzen van herstellmiddelen gebaseerd op de graad van een knoop toont potentie om effectiever het aantal infecties terug te brengen dan de klassieke, homogene toewijzing.

Gemotiveerd door datasets uit de praktijk gaan we verder in op heterogene infectiesnelheden in hoofdstuk 5 en 6. Met het klassieke SIS-model als maatstaf bestuderen we in hoofdstuk 5 de invloed van onafhankelijk identiek verdeelde infectiesnelheden op het gemiddelde percentage geïnfecteerde knooppunten in de metastabiele toestand, waarin het infectiepercentage niet nul en stabiel is. Onze bevindingen tonen aan dat, wanneer de prevalentie hoog is, de heterogeniteit van infectiesnelheden gemiddeld de verspreiding van een virus vertraagt en dat een groter even-orde moment van de infectiesnelheden tot een lager infectiepercentage leidt, maar dat oneven-orde momenten een tegenovergestelde bijdrage leveren. Wanneer echter de prevalentie laag is, d.w.z. de epidemie mogelijk uitsterft of slechts een klein percentage van de populatie infecteert, zien we dat heterogeniteit de kans dat een epidemie breder uitspreidt kan vergroten. Hoofdstuk 6 gaat een stap verder dan hoofdstuk 5: de heterogene infectiesnelheid tussen elk knooppaar is niet langer onafhankelijk identiek verdeeld, maar gecorreleerd aan de graden van elk van de knooppunten. We zien dat, wanneer een epidemie zeer prevalent is, een negatieve correlatie tussen de graden van twee knooppunten en de infectiesnelheid doorgaans gunstig is voor de verspreiding van de epidemie. Wanneer echter de prevalentie laag is, zorgt een positieve correlatie eerder voor een versterkte verspreiding van de epidemie. Onze resultaten in hoofdstuk 5 en hoofdstuk 6 brengen aan het licht, dat hoe een epidemie zich in de praktijk verspreidt ver kan afwijken van wat een klassiek, homogeen model voorspelt, en onthullen de essentiële rol van heterogeniteit in dynamische processen uit de praktijk.

De voornaamste aanpak in dit proefschrift om het SIS-model met heterogene infectie- of herstelsnelheden te bestuderen zijn simulaties waarin de tijd als continu wordt beschouwd. Hierbij zijn we geïnteresseerd in hoe nauwkeurig een gemiddeldeveldbenadering kan zijn. Derhalve onderzoeken we in hoofdstuk 7 de invloed van heterogene infectiesnelheden op de nauwkeurigheid van NIMFA—een geavanceerd gemiddeldeveldbenadering van het SIS-model.

Zowel de hoofdstukken tot nu toe als de meeste studies rond het SIS-model nemen voornamelijk het percentage geïnfecteerde knooppunten in acht en negeren daarbij de besmettingskans van elke individuele knoop in het netwerk. In hoofdstuk 8 kijken we naar de heterogene prestaties van de knooppunten, d.w.z. de kans op infectie of kwetsbaarheid van elke knoop, met als motivatie dat de ranglijst van knoopkwetsbaarheid gegeven de effectieve infectiesnelheid van een SIS-epidemie cruciale informatie voor een netwerkoperator kan zijn om te begrijpen welke knooppunten kwetsbaarder zijn of welke knooppunten moeten worden beschermd. Via zowel een theoretische als numerieke aanpak tonen we aan dat de ranglijst van knoopkwetsbaarheid neigt om significant te veranderen bij een kleine wijziging in de effectieve infectiesnelheid wanneer deze infectiesnelheid kleiner is dan dat deze groter is. Hetzelfde geldt voor wanneer het netwerk een Barabási-Albert-stochastische graaf is dan dat een Erdős-Rényi-stochastische graaf is.

Als een eerste begin probeert dit proefschrift een overzicht te schetsen hoe de prevalentie van een opinie en epidemie kan worden beïnvloed door verschillende vormen

van heterogeniteit, zoals heterogeniteit in dynamische processen (heterogene infectie- of herstelsnelheden) en in netwerktopologie. Dit proefschrift dient ook ter inspiratie voor toekomstig onderzoek. Hoewel verscheidene soorten heterogeniteit worden beschouwd in dit proefschrift, kan de heterogeniteit die men in de praktijk tegenkomt een stuk complexer en onbenut zijn. Zo is bijvoorbeeld de invloed van tijdsveranderlijke heterogeniteit, d.w.z. de infectie- of herstelsnelheden veranderen en passen zich over tijd aan, nog een open vraagstuk. Verder rest de vraag: hoe kunnen we de optimale immuniseringstrategie ontwerpen wanneer verschillende vormen van heterogeniteit in beschouwing worden genomen?



# 1

## INTRODUCTION

*No man is an island,  
entire of itself;  
every man is a piece of the continent,  
a part of the main.*

John Donne

*For never will be found  
two eggs, or two drops of milk, or two leaves, or two animals,  
and in general two things  
so similar that after an accurate inspection,  
a difference cannot be detected.*

Gottfried Leibniz

## 1.1. BACKGROUND

A Chinese proverb says: “One more friend, one more way”. This ancient Chinese wisdom simply reveals the importance of social relationship for humans and their natural tendency to form *networks*. In such networks, we are connected with each other in many different ways, including the family ties, professional connections or any other type of social relations. The rapid development of the Internet, data generation and sharing technology, as well as data storage and computing, has resulted in extreme networking effects connecting thousands of even millions of people in massive online social networks. The immense commercial success and societal impact of online social networks are mainly due to the *processes* that run in such networks, for instance information exchange or various types of interactions. While information exchange on such a large scale undoubtedly has the potential to enrich our lives, not all effects of such processes are desirable. For instance, the discussions have been raised whether online interaction processes in social networks have started to replace physical interactions among people [117]. Furthermore, the so-called *frictionless sharing* [132] of information – the mechanism underlying modern social networks – has led to the emergence of filter bubbles and the fake-news phenomenon. Consequently, understanding the social networks and the processes inside such networks is critical for us to be able to draw maximal benefit from them, while reducing the undesired effects. This, however, is not trivial in view of a significant *complexity* of such networks, characterized by the number of participating entities (in this case, people), the diversity of links connecting these entities and the complicated, *dynamic* interaction and information exchange processes running across these links over time. Challenges of similar nature can also be encountered in other types of *complex networks*, including brain networks [21, 115], metabolic networks [59, 102], computer networks [54], and finance networks [4].

An analysis of a dynamic process on a complex network requires an effective representation of network *topology* and a network process. Under the network topology we consider a set of nodes representing the entities joined in the network and the links representing the presence of a relation or an interaction between each two entities. By connecting the nodes using links, a *graph* can be constructed, that is typically used to represent a network topology. A large number of measures have been proposed to characterize topological features of networks by looking at various graph properties. Centrality measures have been proposed to capture the specific feature of each node. For example, the *degree* [41] of a node tells us about the number of connections the node has. The *betweenness* [41] of a node is the number of shortest paths<sup>1</sup> that pass through the node. Network measures, on the other hand, have been proposed to capture the features at the network level. For example, the *clustering coefficient* [130] measures how nodes in a network tend to cluster together. This is important since in most real-world networks, and in particular in social networks, the nodes tend to create tight groups characterized by a relatively high density of links. The *assortativity* [88, 133] is used to indicate that the nodes being similar to each other<sup>2</sup> connect to each other. For example, in social networks, which show assortativity, nodes tend to connect to nodes with similar degrees.

<sup>1</sup>For every pair of nodes in a network, there exists a shortest path between the nodes such that the number of links that the path passes through is minimized.

<sup>2</sup>In network theory, assortativity is often examined in terms of a node's degree

As another example, technological and biological networks typically show *dissortativity*, as high degree nodes tend to attach to low degree nodes [88]. Finally, network models have been developed to represent networks that show different global topological characteristics. Examples are the Erdős-Rényi (ER) model [38], the Bárabasi-Albert (BA) scale-free model [9] and the small-world model [131]. For example, the degree distribution of a scale-free network follows a power law, which means that only a few nodes have very large degrees while a large number of nodes have small degrees. Such a feature has been reported in many real-world networks and explained by the preferential attachment process (e.g. in reality, people who are already wealthy receive more than those who are not). Also, the small-world network, possessing a large clustering coefficient, captures the small-world phenomenon (e.g. strangers being linked by a short chain of acquaintances).

Beyond network topologies, stochastic models have been developed to describe the real-world dynamic processes on networks, such as the diffusion of information, the spread of epidemics and the cascade of failures. In such models, the state of any node in a network could change over time and the changes are usually caused by local or global interactions with other nodes. Here we briefly introduce two sorts of models: the opinion models and the epidemic models.

The opinion models are used to study how the opinion changes in a population. For example, a person may first like or dislike a product, and then change his/her opinion because of the influence of a friend. Companies producing the product would be highly interested in understanding how consumers influence each other in their opinion and how to accelerate the propagation of the positive opinion towards their products. For example, in the *non-consensus opinion (NCO) model*, the state of a node, representing the opinion of the node, is binary (to like the product, policy, etc. or not), and the state of each node changes with the states of its neighbors: each node adopts the majority opinion of its neighbors and itself. Many opinion models have been developed for different scenarios, such as the voter model [66, 78, 109], the majority rule model [43, 65] and the social impact model [67, 90].

Epidemic models have been developed to describe, e.g. how an epidemic spreads in a social contact network, how a computer virus spreads in a computer network, or how information propagates [52, 57] in an online social network. Basic epidemic models include the *susceptible-infected (SI) model* [5], the *susceptible-infected-susceptible (SIS) model* [91, 95], and the *susceptible-infected-recovered (SIR) model* [13, 112]. Numerous variants of these models [49, 63, 74] have been proposed as well. Taking the SIS model as an example, a node could either be infected or susceptible at any time. An infected node could recover to be susceptible with a given recovery rate, whereas a susceptible node could be infected by each of its infected neighbors with a given infection rate. This model could be used to describe, for example, how an infectious disease spreads in a population: an infected person could be cured after a treatment, whereas a susceptible person could be infected through a physical contact with infected persons. The fraction of the infected nodes in the steady state reveals the prevalence of the disease, information or the adoption of a product or behavior. A key question that has been widely studied is how the network topology properties influence the prevalence. Given an epidemic spreading model that is used to describe the spreading dynamics of epidemics, virus,

behavior or information, immunization problems have been explored with the objective to immunize a small subset of nodes and minimize the scope of the infection.

## 1.2. MISSION AND SCOPE OF THIS THESIS

"In a forest of a hundred thousand trees, no two leaves are alike. And no two journeys along the same path are alike." Either Paulo Coelho<sup>3</sup> in our time or Gottfried Leibniz<sup>4</sup> in 1600s, they both point out the ubiquitousness of heterogeneities: every object in this world is actually different from another, so is the interaction between any two objects.

In the past, the study of dynamic processes started with the homogeneous dynamic models, where the interaction, e.g. the infection rate, between each pair of nodes is the same, and on the homogeneous networks, where each node has the same number of connections with others. However, most real-world networks are heterogeneous in topology. For example, a few influentials in Twitter have many followers whereas the majority has only a few followers. Therefore, researchers started to consider heterogeneous network topology where different nodes may have different number of connections. For example, epidemic models have been studied intensively on heterogeneous networks [79, 93, 134], where the prevalence of epidemic and the epidemic threshold<sup>5</sup> are theoretically and numerically derived via different approximation methods. Lately, a few studies on epidemic spreadings considered directed networks [72], where the infection may happen in only one direction of a pair of nodes.

In addition to the network topology, real-world dynamic processes also manifest heterogeneities. As an example, the infection rate between two individuals, which depends on their contact frequency, has been shown to be heterogeneous, i.e. different for different pairs of individuals. As more real-world data become traceable, the heterogeneity in dynamic processes has further been investigated. Following the above example, a number of recent works on epidemic models [23, 42, 96, 97, 135] have considered the infection or recovery processes with different rates. Preciado et al. [96, 97] discussed how to choose the infection and recovery rates from given discrete sets to let the virus die out. Fu et al. [42] studied only the epidemic threshold when the infection rates depend on the node degrees. Buono et al. [23] considered a specific distribution of infection rates and observed slow epidemic extinction phenomenon. Yang and Zhou [135] gave an edge-based mean-field solution of the epidemic threshold in homogeneous networks with i.i.d. heterogeneous infection rates (following uniform or power-law distribution).

What can be observed from the literature is that heterogeneities in networks have been considered, but mainly partially: either heterogeneous network topologies were investigated in combination with homogeneous processes, or dynamic processes with very specific heterogeneities, such as a specific infection-rate distribution, were investigated on homogeneous networks or undirected heterogeneous networks. We still do

<sup>3</sup>Paulo Coelho de Souza (born 24 August, 1947) is a Brazilian lyricist and novelist.

<sup>4</sup>Gottfried Wilhelm (von) Leibniz (1 July, 1646 – 14 November, 1716) was a German polymath and philosopher who occupies a prominent place in the history of mathematics and the history of philosophy, having developed differential and integral calculus independently of Isaac Newton.

<sup>5</sup>The epidemic will die out if the ratio of the infection rate to the recovery rate is below the epidemic threshold. Hence, the epidemic threshold is also an important indicator to show how vulnerable a network is against the epidemic other than the prevalence.

not know enough about heterogeneous processes running on heterogeneous networks, e.g. how should we model the heterogeneous network topologies and processes more realistic and what is the influence of such heterogeneities on the prevalence of an opinion or an epidemic. Furthermore, most studies have looked only at the prevalence, the average infection probability of an epidemic on a network, but ignored the heterogeneous performance of each node, the infection probability of a node. The heterogeneous performance of the nodes could be significant for designing the optimal immunization strategy.

The first part of the mission of this thesis is to provide more insight in this insufficiently explored case, i.e., the insufficient study on the realistic heterogeneities. In order to make our investigation as relevant to real-world problems as possible, in addition to undirected networks, we particularly consider heterogeneous directed networks in parts of this thesis. While such networks can often be encountered, both opinion dynamics and epidemic spreading have been seldom studied in these network contexts. Taking the epidemic spreading process as an example, we characterize and model the heterogeneous infection rates in a systematic way such that our infection rate models could capture several key features of real-world infection rates. Finally we rely on real-world datasets to evaluate our models.

The second part of the mission is to provide more insight in the influence of heterogeneity on dynamic processes especially on the prevalence of an epidemic or opinion. More specifically, this thesis investigates how heterogeneities influence the prevalence of an opinion or epidemic and how heterogeneities can be used to directly control the epidemic spreading or indirectly inspire the immunization strategies.

The third part of the mission of the thesis is to explore the heterogeneous performance of the nodes in e.g. an epidemic spreading process. Instead of the prevalence, we investigate further the infection probability per node, or in other words, how the ranking of the nodal infection probability would change as the infection rate varies in a broad range and in different types of networks. The ranking of nodal infection probability, which reflects the heterogeneous performance of nodes, can be crucial for network operators to assess which nodes are more vulnerable or should be protected.

### 1.3. OUTLINE OF THIS THESIS

The technical part of the thesis starts with Chapter 2, in which we introduce the network models and the dynamic models studied in this thesis, including the non-consensus opinion (NCO) model and the susceptible-infected-susceptible (SIS) model. Moreover, we also introduce there an individual-based mean field approximation of the SIS model called the N-Intertwined Mean-Field Approximation (NIMFA). The scientific contribution by which we pursue the mission defined in the previous section is organized as illustrated in Fig. 1.1 and can be found in chapters 3 to 8.

Chapter 3 and Chapter 4 consider the heterogeneity in network topology, particularly the directed network. In Chapter 3, we study the NCO model on different types of directed networks and explore how the fraction of nodes with a given opinion changes. Chapter 4 concentrates on the SIS model with correlated heterogeneous recovery rates on directed networks, where the recovery rate of a node is correlated with the number of its in-links (by which other nodes affect this node) and out-links (by which this node

Heterogeneity in topology: <i>Directed networks</i>	Chapter 3 opinion dynamics, directed networks	Heterogeneity in process: <i>Heterogeneous recovery or infection rates</i>
	Chapter 4: epidemic spreading, directed networks, heterogeneous recovery rates	
Chapter 5: epidemic spreading, independent and identically distributed (i.i.d.) infection rates		
Chapter 6: epidemic spreading, correlated infection rates		
Chapter 7: epidemic spreading, accuracy of mean-field both i.i.d. and correlated infection rates		
	Chapter 8: epidemic spreading, ranking of infection probability	<i>Focus on each node</i>

Figure 1.1: The structure of this thesis.

affects others). There, we discuss how to suppress the epidemic spreading by allocating recovery rates.

Chapter 5, 6 and 7 all consider the SIS model with heterogeneous infection rates on different types of undirected networks, and thus the heterogeneity in dynamic processes. In Chapter 5, we start from the independent and identically distributed (i.i.d.) infection rates, whose distributions are modeled based on real-world datasets. We investigate how the average fraction of infected nodes in a network is affected by the distribution, and especially by the variance of the infection rates. Chapter 6 discusses a more complex infection-rate scenario, where the infection rate is correlated with the degrees of the two end nodes of a link, as motivated by our observations in real-world systems. We explore how the prevalence of an epidemic is influenced by such heterogeneous infection rates. In Chapter 5 and Chapter 6, we explore the heterogeneous epidemic spreading on heterogeneous networks via mainly continuous-time simulations, but also via a theoretical approach – the mean-field approximation of the SIS model. In Chapter 7, we evaluate the precision of the mean-field approximation.

All the above chapters, which are devoted to the first and second parts of the mission of this thesis, focus on computing the average state of all nodes, such as the fraction of nodes with a given opinion in the NCO model or the average fraction of infected nodes i.e. the average infection probability of all nodes. In Chapter 8, we shed light on a new direction: instead of the average of all nodes in the classic SIS epidemic spreading, we explore the state of each node. Particularly, we explore if the ranking of the nodal infection probability could change with the effective infection rate, in which range of the effective infection rates and in which types of networks such changes are more dramatic. This chapter serves the third part of the mission of this thesis.

Chapter 9 concludes the thesis, highlights the main contributions and discusses the directions for future work.

## 1.4. LIST OF PUBLICATIONS RELATED TO THE THESIS

The following papers have been completed by the author of this thesis while pursuing the Ph.D. degree at the Delft University of Technology. Those publications directly serving as chapters of the thesis are indicated accordingly.

### JOURNAL

4. **B. Qu** C. Li, P. Van Mieghem and H. Wang, The nodal infection probability in SIS epidemic spreading, *Accepted by Scientific Report*. – [**Chapter 8**]
3. **B. Qu** and H. Wang, SIS Epidemic Spreading with Correlated Heterogeneous Infection Rates, *Physica A Statistical Mechanics & Its Applications*, 2017, 472(23–24):4543–4548. – [**Chapter 6**]
2. **B. Qu** and H. Wang, SIS Epidemic Spreading with Heterogeneous Infection Rates, *IEEE Transactions on Network Science and Engineering*, Issue 99, 2017. – [**Chapter 5**]
1. **B. Qu**, Q. Li, S. Havlin, H.E. Stanley and H. Wang, Nonconsensus opinion model on directed networks, *Physical Review E*, 2014, **90**(5): 052811. – [**Chapter 3**]

### CONFERENCE

2. **B. Qu** and H. Wang, The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading, *Complex Networks 2016*. . – [**Chapter 7**]
1. **B. Qu**, A. Hanjalic, and H. Wang, Heterogeneous Recovery Rates against SIS Epidemics in Directed Networks, *NetGCoop 2014: International Conference on NETWORK Games, CONTROL and OPTimization*. Trento, Italy. – [**Chapter 4**]



# 2

## NETWORK STRUCTURES AND DYNAMIC PROCESSES

*The dynamic processes on complex networks have been developed to represent the information propagation for years. On one hand, complex networks, modeling a large population, have been described by mathematical graph theory. On the other hand, dynamic processes have been studied by various deterministic or compartmental mathematical models. This chapter first introduces the network models and then elaborate the dynamic models investigated through this thesis.*

## 2.1. NETWORK STRUCTURES

**I**N graph theory, a graph is made up of a set  $\mathcal{N}$  of nodes which are connected by links in a set  $\mathcal{L}$ . Such a graph  $G = (\mathcal{N}, \mathcal{L})$  is usually used to represent a complex network. The numbers of nodes and links are then denoted by  $N = |\mathcal{N}|$  and  $L = |\mathcal{L}|$  respectively. Node  $i$  is a neighbor of node  $j$  if there is a link connecting the two nodes. The degree  $d_i$  of node  $i$  is the number of its neighbors. The degree vector is denoted by  $d = [d_1 \ d_2 \ \dots \ d_N]^T$  and the degree of an arbitrary node is denoted by the random variable  $D$ . The basic law for the degree is  $\sum_{i=1}^N d_i = 2L$ , and the average degree  $E[D] = \frac{\sum_{i=1}^N d_i}{N} = \frac{2L}{N}$ . The  $N \times N$  adjacency matrix  $A$ , consisting of elements of  $a_{ij}$  where  $a_{ij} = 1$  if there is a link from node  $i$  to node  $j$  or else  $a_{ij} = 0$ , is a representation of a graph. The largest eigenvalue of the adjacency matrix  $A$  is  $\lambda_1$ , also called the spectral radius. The principal eigenvector  $x_1$  corresponding to the largest eigenvalue  $\lambda_1$  satisfies  $Ax_1 = \lambda_1 x_1$ . The  $i$ -th component of the principal eigenvector is denoted by  $(x_1)_i$ .

### 2.1.1. ERDŐS-RÉNYI RANDOM GRAPHS

The Erdős-Rényi (ER) random graph [38] is one of the most widely-used and well-studied models. In an ER random graph  $G_p(N)$  with  $N$  nodes, each pair of nodes is connected with probability  $p$  independent from every other pair. The distribution of the degree of a random node is binomial:  $\Pr[D = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k}$  and the average degree  $E[D] = (N-1)p$ . For large  $N$  and constant  $E[D]$ , the degree distribution tends [123] to a Poisson distribution:  $\Pr[D = k] = e^{-E[D]} (E[D])^k / k!$ . Moreover, if the link density  $p > p_c = \frac{\ln N}{N}$  for large  $N$ , the graph  $G_p(N)$  is almost surely connected.

### 2.1.2. BARABÁSI-ALBERT RANDOM GRAPHS

Besides the ER random graph, the scale-free model is often used to capture the scale-free degree distribution of the real-world networks such as the Internet [24] and World Wide Web [2]. In this thesis, we consider the SF networks generated by the Barabási-Albert (BA) model and the configuration model. The BA model is first introduced in this subsection and then the configuration model in Section 2.1.3.

The BA network begins with an initial connected network of  $m_0$  nodes. New nodes are added to the network one at a time. Each new node is connected to  $m \leq m_0$  existing nodes with a probability that is proportional to the number of links that the existing nodes already have. The degree distribution of BA random graphs [123] is given by  $\Pr[D = k] = ck^{-3}$  for sufficiently large  $N$ , where  $c = (\sum_{k=m}^{N-1} k^{-3})^{-1}$  and the minimum degree is  $m$ . We set  $m_0 = m + 1$  and start with a full graph with  $m_0$  nodes, then add one new node connecting  $m$  existed nodes at a time until the graph size is expected. Hence, the number of links is  $L = m_0(m_0 - 1)/2 + (N - m_0)m = (N - \frac{m_0}{2})m$  and the average degree is  $E[D] = \frac{2L}{N} = \frac{2N - m_0}{N}m$ , thus approximately  $2m$  in this work.

### 2.1.3. CONFIGURATION MODEL

Given a degree distribution  $\Pr[D = k]$ , the configuration model can be used to build up random graphs. The degree vector  $d = [d_1 \ d_2 \ \dots \ d_N]^T$  is first generated by comparing  $N$  uniformly distributed random numbers  $r_i \in (0, 1)$  with the cumulative distribution function (CDF)  $\Pr[D \leq k]$ :  $d_i = k$  if  $r_i > \Pr[D \leq k - 1]$  and  $r_i \leq \Pr[D \leq k]$ . A sequence

of "stubs" (one end of a link) is then constructed by repeating the index  $i$  of node  $i$   $d_i$  times. Afterwards, the  $2L$  stubs are randomly paired. Hence, the nodes are randomly connected.

## 2.2. DYNAMIC PROCESSES

Dynamic processes, such as the percolation [116], epidemics spreading [15, 28, 33, 69, 111, 121, 128], opinion dynamics [34, 43, 56, 65–67, 90, 109, 118], and cascading failures [26], have been intensively studied in recent years. In this section, the models considered in this thesis are introduced.

### 2.2.1. OPINION DYNAMICS

Treating opinion as a variable allows us to model patterns of opinion formation as a dynamic process on a complex network with nodes as agents and links as interactions between agents. Shao et al. [110] proposed a non-consensus opinion (NCO) model that achieves a steady state in which two opinions can coexist. Their model reveals that when the initial population of a minority opinion is above a certain critical threshold, a large steady-state spanning cluster with a size proportional to the total population is formed [110]. This NCO complex network model belongs to the same universality class as percolation [22, 110, 116], and have received much attention.

#### NCO MODEL

In a NCO model [110] on a single network with  $N$  nodes, each with binary opinions, a fraction  $f$  of nodes has opinion  $\sigma_+$  and a fraction  $1 - f$  has opinion  $\sigma_-$ . The opinions are initially randomly assigned to each node. At each time step, each node adopts the majority opinion, when considering both its own opinion and the opinions of its nearest neighbors (the agent's friends). A node's opinion does not change if there is a tie. Following this opinion formation rule, at each time step the opinion of each node is updated. The updates occur simultaneously and in parallel until a steady state is reached. Note that when the initial fraction  $f$  is above a critical threshold,  $f \equiv f_c$  (even minority), both opinions continue to exist in the final steady state.

Figure 2.1 shows an example of the dynamic process of the NCO model on a small directed network with ten nodes. Here we consider the in-neighbors of a node as the friends influencing the node, and the out-neighbors as the friends influenced by the node. At time  $t = 0$  five nodes are randomly assigned the opinion  $\sigma_+$  (empty circle), and the other five nodes the opinion  $\sigma_-$ . At time  $t = 0$  nodes  $A_1$  and  $A_2$  have opinion  $\sigma_-$  and  $\sigma_+$  respectively but are in a local minority and thus updating it means changing their opinions to  $\sigma_+$  and  $\sigma_-$  respectively. At time  $t = 1$  node  $B$  belongs to a local minority and thus needs updating. At time  $t = 3$  all nodes hold the same opinion as their local majority, and the system has reached a final non-consensus steady state.

### 2.2.2. EPIDEMIC SPREADING

A network node could be infected, become a new infection source and infect other hosts. On the other hand, network nodes are usually equipped with certain recovery resources, so that they can be recovered from the infection in a finite time. The infection and recovery processes above have been described by epidemic models[3, 25, 45, 46, 60, 92]. One

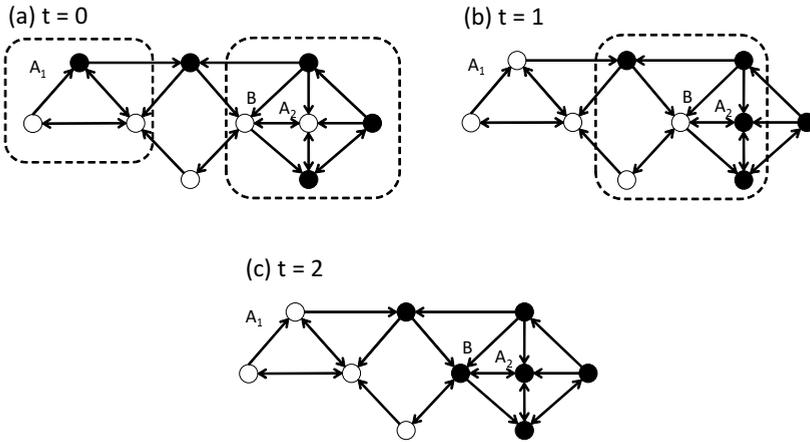


Figure 2.1: Schematic plot of the dynamics of the NCO model on a directed graph with 10 nodes.

of the most widely studied models is the susceptible-infected-susceptible (SIS) model[5, 6, 8, 15, 28, 33, 64, 69, 111, 121, 136].

#### THE SIS MODEL

In the continuous-time Markovian SIS model on a network with  $N$  nodes, the state of a node at any time  $t$  is a Bernoulli random variable  $X_i(t)$ , where  $X_i(t) = 0$  represents that node  $i$  is susceptible and  $X_i(t) = 1$  that node  $i$  is infected. Each infected node infects each of its susceptible neighbors with an infection rate  $\beta$ . The infected node can recover with a curing rate  $\delta$ . Both the infection and recovery process are independent Poisson processes. The ratio  $\tau = \frac{\beta}{\delta}$  is called the effective infection rate. Independent of the effective infection rate  $\tau$ , the infection process dies out in any finite graph after an unrealistically long time, and the corresponding steady state is the absorbing state, i.e. the overall-healthy state. However, if  $\tau$  is larger than the epidemic threshold  $\tau_c$ , there is a non-trivial metastable state, where the average fraction of infected nodes is non-zero and stable for a long time [124]. The vulnerability or the infection probability  $v_{k\infty}$  is defined as the metastable state infection probability of node  $k$ . The average fraction of infected nodes in the metastable state is denoted by  $y_\infty = \frac{1}{N} \sum_{j=1}^N v_{j\infty}$ .

#### THE N-INTERTWINED MEAN-FIELD APPROXIMATION OF THE SIS MODEL

The N-Intertwined Mean-Field Approximation (NIMFA) is one of the most accurate approximation of the SIS model that takes into account the influence of the network topology [121]. The single governing equation for a node  $i$  in NIMFA is

$$\frac{dv_i(t)}{dt} = -\delta v_i(t) + \beta(1 - v_i(t)) \sum_{j=1}^N a_{ij} v_j(t) \quad (2.1)$$

where  $v_i(t)$  is the infection probability of node  $i$  at time  $t$ , and the adjacency matrix element  $a_{ij} = 1$  or 0 denotes if there is a link or not between node  $i$  and node  $j$ . With

$V(t) = [v_1(t) \ v_2(t) \ \cdots \ v_N(t)]^T$ , the matrix evolution equation of NIFMA is

$$\frac{dV(t)}{dt} = (\beta \text{diag}(1 - v_i(t))A - \delta I)V(t) \quad (2.2)$$

where  $A$  is the  $N \times N$  adjacency matrix of the network,  $I$  is the  $N \times N$  identity matrix and  $\text{diag}(v_i(t))$  is the diagonal matrix with elements  $v_1(t), v_2(t), \dots, v_N(t)$ . In the steady state, defined by  $\frac{dV(t)}{dt} = 0$ , or equivalently  $\lim_{t \rightarrow \infty} v_i(t) = v_{i\infty}$  and  $\lim_{t \rightarrow \infty} V(t) = V_\infty$ , we have

$$(\tau \text{diag}(1 - v_{i\infty})A - I)V_\infty = 0 \quad (2.3)$$

Given the network and the effective infection rate  $\tau$ , we can numerically compute the infection probability  $v_{i\infty}$  as a function of the effective infection rate  $\tau$  for each node  $i$  by solving (2.3).

Furthermore, the NIMFA epidemic threshold  $\tau_c^{(1)} = \frac{1}{\lambda_1}$ , where  $\lambda_1$  is the largest eigenvalue of the adjacency matrix  $A$ , is a lower bound of the exact epidemic threshold  $\tau_c$ , i.e.  $\tau_c^{(1)} < \tau_c$ . The Laurent series of the steady-state infection probability is given by [123]

$$v_{i\infty}(\tau) = 1 + \sum_{m=1}^{\infty} \eta_m(i) \tau^{-m} \quad (2.4)$$

possesses the coefficients  $\eta_1(i) = -\frac{1}{d_i}$  and

$$\eta_2(i) = \frac{1}{d_i^2} \left( 1 - \sum_{j=1}^N \frac{a_{ij}}{d_j} \right) \quad (2.5)$$

and for  $m \geq 2$ , the coefficients obey the recursion

$$\eta_{m+1}(i) = -\frac{1}{d_i} \eta_m(i) \left( 1 - \sum_{j=1}^N \frac{a_{ij}}{d_j} \right) - \frac{1}{d_i} \sum_{k=2}^m \eta_{m+1-k}(i) \sum_{j=1}^N a_{ij} \eta_k(j)$$

#### THE CONTINUOUS-TIME SIMULATION OF THE EXACT SIS MODEL

In order to study the effect of the various heterogeneities on the virus spread, we further develop a continuous-time simulator, which was firstly proposed by van de Bovenkamp and described in detail in [69]. A discrete-time simulation could well approximate a continuous process if a small time bin to sample the continuous process is selected so that within each time bin, no multiple events occur. A heterogeneous SIS model allows different as well large infection or recovery rates, which requires even smaller time bin size and challenges the precision of a discrete-time simulation. Hence, we implement the precise continuous-time simulations.



# 3

## OPINION DYNAMICS ON DIRECTED NETWORKS

*Dynamic social opinion models have been widely studied on undirected networks, and most of them are based on spin interaction models that produce a consensus. In reality, however, many networks such as Twitter and the World Wide Web are directed and are composed of both unidirectional and bidirectional links. Moreover, from choosing a coffee brand to deciding who to vote for in an election, two or more competing opinions often co-exist. In response to this ubiquity of directed networks and the coexistence of two or more opinions in decision-making situations, we study a non-consensus opinion model introduced by Shao et al. on directed networks. We define directionality  $\xi$  as the percentage of unidirectional links in a network, and we use the linear correlation coefficient  $\rho$  between the indegree and outdegree of a node to quantify the relation between the indegree and outdegree. We introduce two degree-preserving rewiring approaches which allow us to construct directed networks that can have a broad range of possible combinations of directionality  $\xi$  and linear correlation coefficient  $\rho$  and to study how  $\xi$  and  $\rho$  impact opinion competitions. We find that, as the directionality  $\xi$  or the indegree and outdegree correlation  $\rho$  increases, the majority opinion becomes more dominant and the minority opinion's ability to survive is lowered.*

---

This chapter have been published as: **Bo Qu, Qian Li, Shlomo Havlin, Nonconsensus opinion model on directed networks, Physical Review E, 2014, 90(5): 052811 [99].**

### 3.1. INTRODUCTION

**T**HE study of social networks is one of the most important applications of graph theory. Social scientists began refining the empirical study of networks in the 1970s, and many of the mathematical and physical tools currently used in network science were originally developed by them [86]. Social network science has been used to understand the diffusion of innovations, news, and rumors as well as the spread of disease and health-related human behavior [20, 31, 47, 51, 61, 121, 128]. The decades-old hot topic of opinion dynamics continues to be a central focus among researchers attempting to understand the opinion formation process. Although it may seem that treating opinion as a variable or a set of variables is too reductive and the complexity of human behavior makes such an approach inappropriate, often human decisions are in response to limited options: to buy or not to buy, to choose Windows or Linux, to buy Procter & Gamble or Unilever, to vote for the Republican or the Democrat.

Treating opinion as a variable allows us to model patterns of opinion formation as a dynamic process on a complex network with nodes as agents and links as interactions between agents. Although the behavior dynamics of human opinion are complex, statistical physics can be used to describe the “opinion states” within a population and also the underlying processes that control any transitions between them [1, 14, 19, 27, 35, 44, 92]. Over the past decade numerous opinion models have combined complex network theory and statistical physics. Examples include the Sznajd model [118], the voter model [66, 78, 109], the majority rule model [43, 65], the social impact model [67, 90], and the bounded confidence model [34, 56]. All of these models ultimately produce a consensus state in which all agents share the same opinion. In most real-world scenarios, however, the final result is not consensus but the coexistence of at least two differing opinions.

Shao et al. [110] proposed a non-consensus opinion (NCO) model that achieves a steady state in which two opinions can coexist. Their model reveals that when the initial population of a minority opinion is above a certain critical threshold, a large steady-state spanning cluster with a size proportional to the total population is formed [110]. This NCO complex network model belongs to the same universality class as percolation [22, 110], and have received much attention. Among the variants are a NCO model with inflexible contrarians [75] and a NCO model on coupled networks [12, 76].

To date the model has not been applied on directed networks. Directed networks are important because many real-world networks, e.g., Twitter, Facebook, and email networks, are directed [72]. In contrast to undirected networks, directed networks contain unidirectional links. In opinion models, a unidirectional link between two nodes indicates that the influence passing between the two nodes is one-way. A real-world example might be a popular singer who influences the opinions the fans hold, but the fans do not influence the singer’s opinion. In contrast, bidirectional links occur when the influence between two agents is both ways. Real-world unidirectional links are ubiquitous and strongly influence opinion formation, i.e., widespread one-way influence has a powerful effect on opinion dynamics within a society.

Our goal here is to examine how the NCO model behaves on directed networks. We compare the results of different networks in which we vary the proportion of unidirectional links. We also measure the influence of asymmetry between indegree and outdegree. We find that when the indegree and outdegree of each node are the same, an

increase in the number of unidirectional links helps the majority opinion spread and when the fraction of unidirectional links is at a certain level, increasing the asymmetry between indegree and outdegree increases the minority opinion's ability to survive. We also observe that changing the proportion of the unidirectional links or the relationship between the indegree and outdegree of the nodes causes phase transitions.

## 3.2. BASIC DEFINITIONS AND NOTATIONS

### 3.2.1. THE DIRECTIONALITY $\xi$ AND INDEGREE OUTDEGREE CORRELATION

$\rho$

To quantitatively measure the one-way influence in a network, we define the directionality  $\xi$  as the ratio between unidirectional links and all links. The directionality is  $\xi = L_{\text{unidirectional}}/L$ , where the normalization  $L = L_{\text{unidirectional}} + 2L_{\text{bidirectional}}$ , because a bidirectional link can be considered as two unidirectional links. Because we want to determine how much one-way influence affects the NCO model, we consider as a variable the fraction of one-way links  $\xi$ , where  $\xi = 0$  represents undirected networks. Although the sum of indegree and the sum of outdegree are equal in a directed network, the indegree and outdegree of a single node are usually not the same. To quantify the possible difference between the node's indegree and outdegree, we use the linear correlation coefficient  $\rho$  between them,

$$\rho = \frac{\sum_{i=1}^N (d_{i,\text{in}} - E[D])(d_{i,\text{out}} - E[D])}{\sqrt{\sum_{i=1}^N (d_{i,\text{in}} - E[D])^2} \sqrt{\sum_{i=1}^N (d_{i,\text{out}} - E[D])^2}} \quad (3.1)$$

where  $d_{i,\text{in}}$  and  $d_{i,\text{out}}$  are the indegree and outdegree of node  $i$  respectively. The average degree  $E[D]$  is the same for both indegree and outdegree. Note that when  $\rho = 1$  the indegree is linearly dependent on the outdegree for all nodes, and when  $\rho = 0$  the indegree and outdegree are independent of each other. In this paper we confine ourselves to the case in which the indegree and outdegree follow the same distribution. In this case,  $\rho = 1$  implies that  $d_{i,\text{in}} = d_{i,\text{out}}$  holds for every node  $i$ .

## 3.3. ALGORITHM DESCRIPTION

Inspired by earlier research on directed networks [76, 89, 106, 108, 137], we propose two algorithms to construct directed networks. One is a rewiring algorithm that can be applied to any existing undirected network to obtain a directed network with any given directionality but each node has the same indegree and outdegree as the original undirected network. The other constructs directed networks with a given directionality and indegree-outdegree correlation, and with the same given indegree and outdegree distribution. Note that all networks considered in this paper contain neither self-loops nor multiple links in one direction between two nodes.

### 3.3.1. DIRECTIONALITY-INCREASING REWIRING (DIR)

Here we introduce a rewiring approach that changes the directionality but does not change the indegree and outdegree of any node. It was first proposed in Ref. [127], and

also employed by Ref. [72]. Here we improve it to gradually increase the directionality, via a technique we call directionality-increasing rewiring (DIR).

Many undirected network models with various properties have been designed. Examples include the Erdős-Rényi model [38], the Barabasi-Albert scale-free model [9], and the small-world model [131]. If the links of an undirected graph are considered bidirectional, for an arbitrary undirected graph the indegree and outdegree correlation will be  $\rho = 1$ . Figure 3.1 demonstrates an approach that changes the directionality but does not change the indegree and outdegree of any node nor  $\rho$ . We randomly choose two bidirectional links connecting four nodes and treat them as four unidirectional links. Note that this differs from the technique presented in Ref. [72] in that we choose two bidirectional links instead of two random links that may also contain unidirectional links so that the directionality increases after each step. Then we choose two unidirectional links, one from each bidirectional link, and rewire them as follows. For both unidirectional links the head of one link is replaced with the head of the other. If this rewiring introduces multiple links in any direction between any two nodes, we discard it and randomly choose two other bidirectional links. We can increase the number of unidirectional links by repeating the rewiring step and increasing the directionality in each step. The directionality  $\xi$  can be varied from 0 to 1. In general, DIR can be applied to any directed network to further increase its directionality.

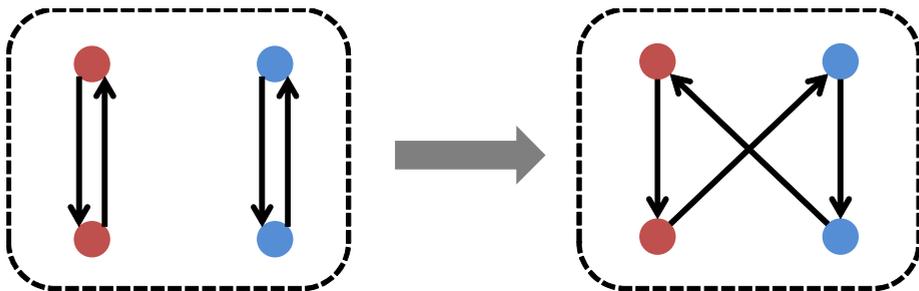


Figure 3.1: (Color online) Directionality-increasing rewiring (DIR)

### 3.3.2. CONSTRUCTING AN ASYMMETRIC INDEGREE AND OUTDEGREE NETWORK AND REWIRING IT TO DECREASE ITS DIRECTIONALITY (ANC-DDR)

We have shown how to obtain a desired directionality  $\xi$  when the indegree and outdegree correlation is  $\rho = 1$ . We further propose an algorithm to construct a network with a given combination of  $\xi$  and  $\rho$ , where  $\rho \neq 1$ . Inspired by the work presented in Ref. [108], which focuses on generating directed scale free (SF) networks with correlated indegree and outdegree sequences, we extend it to a scenario in which the indegree and outdegree sequences follow a distribution that is arbitrary but the same, and we control not just the correlation between the indegree and outdegree but also the directionality, which was ignored in Ref. [108]. We generate an indegree sequence (following a Poisson distribution or power law) and a null outdegree sequence. We then copy a fraction  $\rho$  of the indegree

sequence to the outdegree sequence, and shuffle the fraction  $1 - \rho$  of the indegree sequence as the rest of the outdegree sequence. We thus create an outdegree sequence, a fraction  $\rho$  of which is identical to the corresponding part of the indegree sequence and a fraction  $1 - \rho$  of which is independent of the indegree sequence. After randomly connecting all nodes (given their indegree and outdegree), as in the configuration model [89], we obtain a network with a directionality<sup>1</sup>  $\xi \approx 1$  and an indegree and outdegree correlation close to  $\rho$ . Note that we can further control the indegree and outdegree correlation in a small range close to  $\rho$  by discarding networks with indegree and outdegree correlations outside the expected range. This enables us to construct a network with the indegree and outdegree correlation  $\rho$  ( $0 \lesssim \rho \leq 1$ ), a technique we call asymmetric indegree-outdegree network constructing (ANC).

We use the following rewiring steps to further tune the directionality without changing the indegree and outdegree of each node or the indegree and outdegree correlation  $\rho$ . The goal is to decrease the directionality by repeatedly rewiring two unidirectional links into one bidirectional link. In each step, we choose four nodes linked with at least three directed links as shown on the top half of Fig. 3.2(a). We rewire these three links to the positions shown at the bottom of Fig. 3.2(a). If this rewiring introduces multiple links between any two nodes in any direction we discard the rewiring, select four new nodes, and repeat the step<sup>2</sup>. This rewiring produces at least one more bidirectional link and thus decreases the directionality. We call this procedure directionality-decreasing rewiring (DDR). We combine DDR with ANC and call the entire algorithm ANC-DDR. It seems that ANC-DDR may introduce disconnected components. However, we will see later in Section 3.5 that, the networks generated by all the algorithms are well connected, i.e. almost all the nodes are included in the largest component.

Using ANC we can construct a network with a specified indegree and outdegree correlation  $\rho$ , where the indegree and the outdegree follow the same given distribution and, using DDR, we can change the directionality  $\xi$  in a range dependent on the given  $\rho$  without changing the indegree and outdegree. The range within which we can tune  $\xi \in [\xi_{\min}, 1]$  depends on the given  $\rho$ . For example, for binomial networks<sup>3</sup>,  $\xi$  can be changed from 0 to 1 when  $\rho = 1$ , but the minimum value of  $\xi$  must be approximately 0.3 and any smaller  $\xi$  value is disallowed when  $\rho = 0$ . We explore the relation between the minimal possible directionality  $\xi$  and a given indegree and outdegree correlation  $\rho$  first via numerical simulations<sup>4</sup> in both binomial and SF networks<sup>5</sup>. Figure 3.2(b) shows the linear relationship in both types of network. Binomial networks are characterized by a Poisson degree distribution which is the same as ER random graphs.

For any network constructed using ANC-DDR with an arbitrary given degree distribution  $\Pr[D = k]$  (where the distribution is same in both indegree and outdegree), we can analytically prove the relationship between the minimal possible directionality  $\xi_{\min}$

<sup>1</sup> $E[\xi] = 1 - E[D]^2 N / (N - 1)^2, \lim_{N \rightarrow +\infty} E[\xi] = 1$ .

<sup>2</sup>An efficient rewiring program is available upon request

<sup>3</sup>*Binomial networks* are directed networks with the same Poissonian indegree and outdegree distributions.

<sup>4</sup>In each realization of the simulations, we apply DDR repeatedly on the network constructed by ANC until the four-node structure in Fig. 3.2(a) cannot be found after a number  $M$  of consecutive attempts, then the directionality  $\xi$  is considered the minimal directionality  $\xi_{\min}$  corresponding to the given  $\rho$ . For each given  $\rho$ , we perform 100 realizations and calculate the average of the minimal directionality  $\xi_{\min}$ .

<sup>5</sup>*SF networks* are directed networks whose indegree and outdegree distributions follow the same power law.

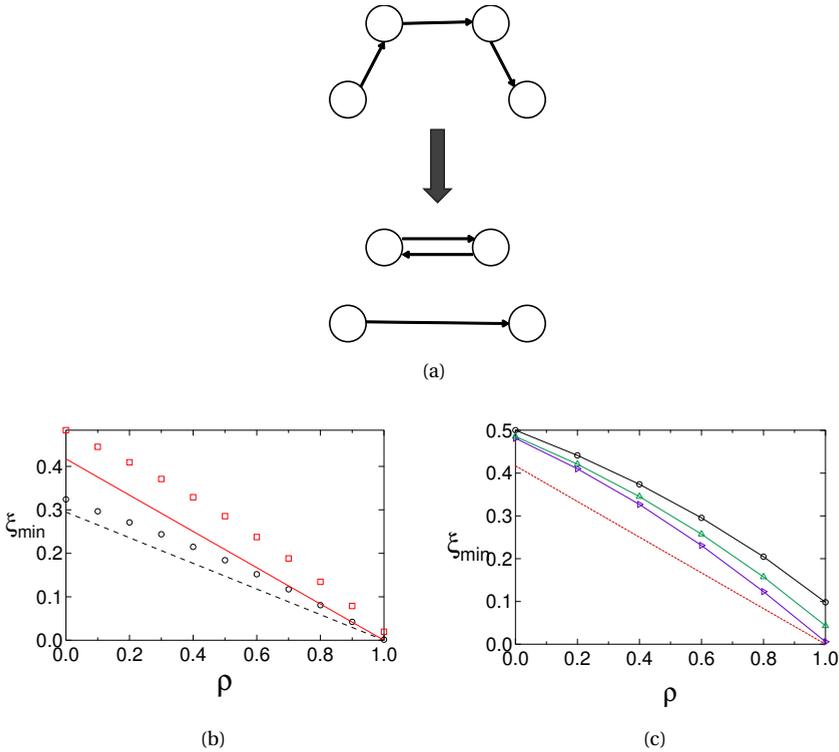


Figure 3.2: (Color online) (a) The degree-preserving rewiring for decreasing the directionality. (b) Plot of the minimal directionality  $\xi_{\min}$  obtained by simulating ANC-DDR, for binomial networks ( $\circ$ ,  $E[D] = 4$ ,  $10^5$  nodes) and SF networks ( $\square$ ,  $\lambda = 2.63$ ,  $10^5$  nodes) with 100 realizations ( $M = E[D]N$ ), and the theoretical minimum possible directionality  $\xi_{\min}$ , Equation (3.2), for binomial networks (the solid line) and for SF networks (the dash line) as a function of the indegree and outdegree correlation  $\rho$ . (c) Plot of the minimal directionality  $\xi_{\min}$  obtained by simulating ANC-DDR with different values of the attempts  $M$ :  $0.01 * M_0$  ( $\circ$ ),  $0.1 * M_0$  ( $\triangle$ ),  $100 * M_0$  ( $\square$ ), where  $M_0 = E[D]N$ , and the theoretical minimum possible directionality  $\xi_{\min}$  (the dash line) for SF networks. All the results are the averages of 1000 realizations.

and the indegree outdegree correlation  $\rho$ ,

$$E[\xi_{\min}] = \frac{1-\rho}{E[D]} \sum_{k=0}^{N-1} k \Pr[d=k] \left( \sum_{i=0}^k \Pr[d=i] - \sum_{i=k}^{N-1} \Pr[d=i] \right). \quad (3.2)$$

The proof of (3.2) is given as follows:

**Proof.** Given the indegree and outdegree of each node, the minimum directionality can be reached if all the unidirectional links of each node are either indegree links or outdegree links but not both, because unidirectional indegree links and outdegree links of a node may form bidirectional links by rewiring so that the directionality  $\xi$  is further reduced. This means that the minimum directionality can be reached if an arbitrary node  $i$  has only  $|d_{i,\text{in}} - d_{i,\text{out}}|$  unidirectional indegree links or outdegree links but not both, where  $d_{i,\text{in}}$  and  $d_{i,\text{out}}$  represent the indegree and outdegree links of node  $i$ , respectively. Hence the minimum possible directionality, given the number of indegree and outdegree links of each node, is

$$\xi_{\min} = \frac{\sum_{i=1}^N |d_{i,\text{in}} - d_{i,\text{out}}|}{\sum_{i=1}^N (d_{i,\text{in}} + d_{i,\text{out}})}. \quad (3.3)$$

We denote the indegree and outdegree sequences by  $S_{\text{in}} = \{d_{i,\text{in}} | i = 1, 2, \dots, N\}$  and  $S_{\text{out}} = \{d_{i,\text{out}} | i = 1, 2, \dots, N\}$  with the same length  $N$ . The indegree of each node  $d_{i,\text{in}}$  is independent and follows the distribution  $\Pr[D = k]$  with the mean  $E[D]$ . In order to introduce the indegree and outdegree correlation,  $S_{\text{out}}$  is constructed from  $S_{\text{in}}$  as follows: a fraction  $\rho$  of the elements in  $S_{\text{out}}$  equals that in  $S_{\text{in}}$  ( $d_{i,\text{out}} = d_{i,\text{in}}$ , for  $i = 1, 2, \dots, \rho N$ , without loss of generality, we assume  $\rho N$  is an integer), while a fraction  $1 - \rho$  of  $S_{\text{out}}$  is obtained by copying and shuffling the rest of  $S_{\text{in}}$ , such that for  $i > \rho N$  and large  $N$ ,  $d_{i,\text{in}}$  and  $d_{i,\text{out}}$  are independent but follow the same distribution  $\Pr[D = k]$ . Hence,

$$\begin{aligned} E(\xi_{\min}) &= E\left(\frac{\sum_{i=\rho N+1}^N |d_{i,\text{in}} - d_{i,\text{out}}|}{\sum_{i=1}^N (d_{i,\text{in}} + d_{i,\text{out}})}\right) \\ &= (1-\rho)E\left(\frac{N|D_{\text{in}} - D_{\text{out}}|}{\sum_{i=1}^N (D_{i,\text{in}} + D_{i,\text{out}})}\right) \\ &= (1-\rho)E[\xi_{\min,\rho=0}], \end{aligned} \quad (3.4)$$

where  $D_{\text{in}}$  and  $D_{\text{out}}$  are independent random variables following the same probability distribution  $\Pr[D = k]$ , and  $\xi_{\min,\rho=0}$  indicates the value of  $\xi_{\min}$  when  $\rho = 0$ .

We then consider the case when  $\rho = 0$ , i.e.,

$$E(\xi_{\min,\rho=0}) = \frac{E[\text{Max}(D_{\text{in}}, D_{\text{out}})] - E[\text{Min}(D_{\text{in}}, D_{\text{out}})]}{2E[D]}, \quad (3.5)$$

where  $\text{Max}(\dots)$  and  $\text{Min}(\dots)$  are the maximum and minimum functions, respectively.

The minimum of random variables  $D_{\text{in}}, D_{\text{out}}$  has the distribution,

$$\begin{aligned} \Pr[\text{Min}(D_{\text{in}}, D_{\text{out}}) = k] &= \Pr[D_{\text{in}} = k] \Pr[D_{\text{out}} \geq k] + \Pr[D_{\text{out}} = k] \Pr[D_{\text{in}} \geq k] \\ &= 2\Pr[D = k] \sum_{i=k}^{N-1} \Pr[D = i] \end{aligned} \quad (3.6)$$

when the two random variables are independent. In the same way, we have

$$\Pr[\text{Max}(D_{\text{in}}, D_{\text{out}}) = k] = 2\Pr[D = k] \sum_{i=1}^k \Pr[D = i] \quad (3.7)$$

Hence,

$$\mathbb{E}[\xi_{\min, \rho=0}] = \frac{1}{\mathbb{E}[D]} \sum_{k=0}^{N-1} k \Pr[D = k] \left( \sum_{i=0}^k \Pr[D = i] - \sum_{i=k}^{N-1} \Pr[D = i] \right). \quad (3.8)$$

Combining 3.5 and 3.4, we have

$$\mathbb{E}[\xi_{\min}] = \frac{1-\rho}{\mathbb{E}[D]} \sum_{k=0}^{N-1} k \Pr[D = k] \left( \sum_{i=0}^k \Pr[D = i] - \sum_{i=k}^{N-1} \Pr[D = i] \right). \quad (3.9)$$

■

Because of the finite number  $M$  of attempts and the random selection process determining the four-node structure, the  $\xi_{\min}$  obtained using simulations is slightly larger than the theoretical  $\xi_{\min}$ . To further understand the gap between the simulation and theoretical results concerning the relationship between  $\xi_{\min}$  and  $\rho$ , we plot the minimal directionality  $\xi_{\min}$  obtained by simulating ANC-DDR with different values of the attempts  $M$  in Fig. 3.2(c). We can see that, as  $M$  increases, the simulation results becomes closer to the theoretical results, but the simulation results seem not to converge to the theoretical results. Actually, we find our algorithm may end up at the network in which we cannot further rewire the links to lower the directionality even with exhaustive searching, though the directionality of the network is still larger than the theoretical minimal. According to the proof of (3.2), the theoretical minimal possible directionality can be reached if and only if all the unidirectional links of each node are either indegree links or out degree links. However, such structure patterns are mostly unreachable. That's to say, the random process, DDR, almost certainly leads the network to the un-rewirable structure patterns which are not the ones inducing the theoretical minimal directionality.

The computational complexity of our algorithm comes mainly from DDR. Here, we will discuss the total rewiring attempts to glimpse the worst case computational complexity of DDR. For a network constructed by ANC, with  $N$  nodes, the average degree  $\mathbb{E}[D]$  and directionality  $\xi = 1$ , if we want to rewire it to a desirable directionality  $\xi_d$  ( $\leq \xi_{\min}$ ), we have to successfully rewire  $\mathbb{E}[D]N(1-\xi_d)$  times. This is the maximal possible number of successful rewiring steps since the minimal possible directionality is mostly unreachable. In order to obtain one successful rewiring step, in the worst case, we carry out  $M$  rewiring attempts. Hence, in the worst case, the number of the total rewiring attempts including successful and unsuccessful ones is  $M^{\mathbb{E}[D]N(1-\xi_d)}$ .

Although using ANC-DDR we can construct a network with a given  $\rho$  and a given  $\xi$  within a corresponding range to  $\rho$ , we cannot obtain a network with the directionality  $\xi$  close to 0 by DDR. We thus apply DIR to undirected network models in order to generate directed networks with a directionality ranging over  $[0, 1]$ , but with the given indegree and outdegree correlation  $\rho = 1$ , to understand the effect of directionality on opinion

competitions. We then use ANC-DDR to generate directed networks with a given indegree and outdegree distribution and correlation, and a given directionality, to explore the effect of both  $\xi$  and  $\rho$  on the opinion model.

### 3.4. THE INFLUENCE OF THE DIRECTIONALITY

In order to examine how the directionality  $\xi$  influences the NCO model, we apply DIR to undirected network models to generate directed binomial networks, SF networks, and random regular (RR) networks<sup>6</sup> [18] with directionality ranging over  $[0, 1]$ . The NCO model is further simulated on each directed network instance. All simulation results are the average of  $10^3$  networks with  $N = 10^5$  nodes and  $E[D] = 4$ .

We use  $S_1$  to denote the size of the largest  $\sigma_+$  cluster in the steady state (where  $\sigma_+$  is the initial opinion randomly assigned to a fraction  $f$  of nodes) and  $S_2$  to denote the size of the second largest cluster. For all three types of networks, we plot  $s_1 \equiv S_1/N$  and  $s_2 \equiv S_2/N$  as a function of  $f$  for different values of the directionality  $\xi$  in Fig. 3.3(a), 3.3(b), and 3.3(c). Additionally, we plot the number of iterations to the steady state (NOI) as a function of  $f$  for RR and binomial networks in Fig. 3.3(a) and 3.3(b) respectively. Note that, depending on the value of  $\xi$ , there is a critical threshold  $f \equiv f_c$  above which there is a giant steady-state component of opinion  $\sigma_+$ . The peak of  $s_2$  indicates the existence of a second-order phase transition, where  $s_1$  is the order parameter and  $f$  is the control parameter. Note that as the value of  $\xi$  increases, in all networks  $f_c$  shifts to the right, a shift observable from the shift of the peak of  $s_2$ . Moreover, the first peak of NOI also shifts to the right as  $\xi$  increases, and coincide with the peak of the second largest cluster representing the critical  $f_c$  for the minority to emerge as a giant component. The second peak occurs due to symmetry at  $1 - f_c$ . In RR networks we lose the peak of  $s_2$  when the directionality  $\xi$  is close to 1, which suggests the disappearance of the second order phase transition. The sharp jump of  $s_1$  around  $f = 0.5$  also indicates the appearance of an abrupt phase transition. When these networks contain an increasing one-way influence (increasing directionality), in all cases the minority opinion will need a greater number of initial supporters if they are to survive when the steady state is reached.

To further understand this change we consider two extreme cases,  $\xi = 0$  and  $\xi = 1$ . In the former, an agent influences only those who can influence the agent in return. In the latter, an agent influences only those who cannot influence the agent in return. This latter case allows a much more rapid spread of opinions, each agent interacts with a larger number of agents, each has in-neighbors as well as out-neighbors, and the opinion is diffused over a wider area. Note that both the majority and minority opinions can benefit from this wider diffusion, but there is a higher risk that the minority opinion will be devoured at some point. This is the case because the bidirectional link connecting two minority opinion agents benefits the minority opinion—the two agents can encourage each other to keep the minority opinion. When rewiring this kind of link there is a higher probability that the two agents will interact with the majority opinion and thus a higher probability that their opinion will be changed to the majority opinion. Thus rewiring makes it more difficult for the minority opinion to form a stable structure.

<sup>6</sup>In this chapter, *random regular (RR) networks* are directed networks in which the indegrees of all nodes and outdegrees of all nodes are the same and the nodes are randomly connected.

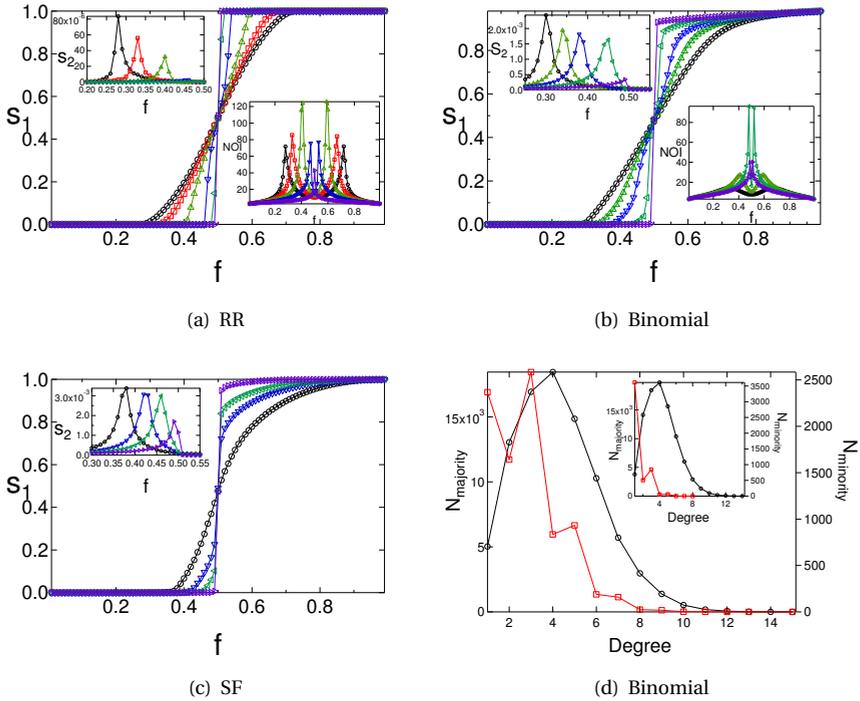


Figure 3.3: (Color online) Plot of the normalized largest cluster  $s_1$  of opinion  $\sigma_+$  as a function of the initial fraction  $f$  for different values of the directionality  $\xi$ : 0( $\circ$ ), 0.2( $\square$ ), 0.4( $\triangle$ ), 0.6( $\nabla$ ), 0.8( $\triangleleft$ ), 1.0( $\triangleright$ ), and for different networks with  $N = 10^5$  nodes and  $E[D] = 4$ : (a) RR, (b) binomial, (c) SF ( $\lambda = 2.63$ ). In the upper insets of (a), (b) and (c), we plot  $s_2$  as a function of  $f$  with the same symbols and for the same networks as in the main figure. In the lower insets of (a) and (b), we plot the number of iterations to the steady state NOI as a function of  $f$ . (d) Plot of the degree distribution of the nodes which keep the majority ( $\circ$ ) and the minority opinion ( $\square$ ) in binomial networks (also with  $N = 10^5$  nodes and  $E[D] = 4$ ), when the directionality  $\xi = 0.0$  (the main figure) and  $\xi = 1.0$  (the inset). All results are based on averaging 1000 realizations.

As directionality  $\xi$  increases, it is easier for minority opinion agents to keep their minority opinion if they have fewer neighbors. Figure 3.3(d) plots the degree distributions (in which the indegree and outdegree follow the same distribution) of the minority-opinion nodes and majority-opinion nodes respectively in the steady state at the critical threshold  $f \equiv f_c$  when the directionality is  $\xi = 0.0$  and  $\xi = 1.0$ . Note that the degrees of most of the minority-opinion nodes that keep their minority opinion are equal to 1, 2, or 3. Minority-opinion nodes with a degree larger than 3 can keep their minority opinion when  $\xi = 0.0$ , but seldom when  $\xi = 1.0$ , i.e., as the value of  $\xi$  increases, the number of nodes following the majority opinion increases, and only low-degree nodes are able to keep the minority opinion. Our results indicate that, a lower directionality helps the existence of the minority opinion, which can be desirable e.g. when the society wants to have different opinions coexisting to inspire or balance each other. We could decrease the directionality of social contact networks by encouraging mutual social interactions e.g. between friends and family members. Moreover, the minority opinion is likely hold by individuals with few social contacts.

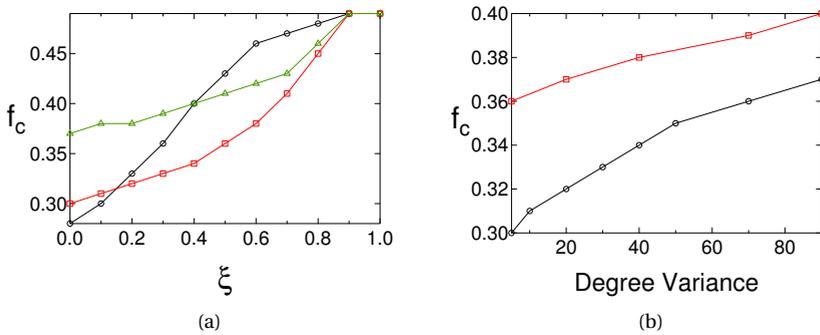


Figure 3.4: (Color online) (a) Plot of the critical threshold  $f_c$  as a function of the directionality  $\xi$  for different networks with  $N = 10^5$  nodes and  $E[D] = 4$ : RR( $\circ$ ), binomial( $\square$ ) and SF( $\triangle$ ) ( $\lambda = 2.63$ ). All results are based on averaging 1000 realizations. (b) Plot of the critical threshold  $f_c$  as a function of the variance of the degree sequence of the networks ( $N = 10^5$  nodes and  $E[D] = 4$ ) with different values of the directionality:  $\xi = 0$ ( $\circ$ ) and  $\xi = 0.5$ ( $\square$ ). All results are based on averaging 100 realizations.

It has been shown that network topology may significantly influence such dynamic processes in networks as epidemics or cascading failures [17, 62, 80, 85, 92]. We thus compare the critical threshold  $f_c$  on directed binomial, RR, and SF networks in which the indegree and outdegree (i) follow the same binomial distribution, (ii) are a constant, and (iii) follow a power-law distribution. Figure 3.4(a) shows that, as the directionality  $\xi$  increases, the critical threshold  $f_c$  of the RR networks increases more rapidly than the others. As stated above, as  $\xi$  increases, only nodes with degrees less than 4, the average degree, are likely to keep the minority opinion, and in RR networks all nodal degrees are 4. Figure 3.4(a) also shows that the existence of hubs (extremely high-degree nodes) in SF networks causes them, at  $\xi = 0$ , to have a much higher critical threshold  $f_c$  than the others, and that the critical threshold in binomial networks is slightly larger than the critical threshold in RR networks. The existence of hubs benefits the majority-opinion

nodes because the probability that an agent with many friends (i.e., a hub) will follow the majority opinion and influence many others is high. They thus strongly contribute to the diffusion of the majority opinion. This phenomenon suggests that public celebrities, i.e. hubs in social networks, tend to help the propagation of the majority opinion.

Reference [104, 107] describes how a second-order phase transition becomes first-order and the critical threshold is higher when the average degree increases. When the average degree is the same, the variance of the degree sequence turns out to be the key factor influencing the critical threshold. In fact, we find that in the networks with the same average degree, the larger the variance of the degree sequence, the larger will be its critical threshold. This is the case because networks with a wider degree variance are more likely to have majority-opinion hubs that can influence many other agents. Figure 3.4(b) shows simulation results that support this behavior. Note that as the variance of the degree sequence increases, the critical threshold increases. To change the variance in these simulations we select a SF network with an average degree  $E[D] = 4$ , randomly remove an existing link, and randomly add a link between two nodes previously unconnected. As we remove and add links repeatedly, the variance of the degree sequence decreases and we stop at an expected variance. To obtain the specified directionality, we apply DIR on the networks. This gives us a wide range of degree variance, which allows us to study the relationship between the variance and critical threshold  $f_c$ .

### 3.5. THE INFLUENCE OF INDEGREE AND OUTDEGREE ASYMMETRY

We have discussed how the critical threshold  $f_c$  increases as the directionality increases in networks in which the indegree and outdegree are the same for each node. The number of in-neighbors and out-neighbors of nodes in real-world networks often differ, however. We mentioned above how a popular singer can influence many people and not be influenced in return. The social network of the singer has many more out-neighbors than in-neighbors. Because this real-world phenomenon is so ubiquitous, we now examine how different correlations between the indegree and outdegree affect opinion competition.

In Section 3.3.2 we use ANC-DDR to construct a network with an arbitrary but identical indegree and outdegree distribution, together with a given combination of the directionality  $\xi$  and the linear correlation coefficient  $\rho$  between the indegree and outdegree. We perform simulations to study the influence of both the directionality  $\xi$  and the correlation coefficient  $\rho$  on the critical threshold  $f_c$ . Figures 3.5(a) and 3.5(b) show that, given the directionality, the critical threshold increases for binomial and SF networks, respectively, as the indegree and outdegree correlation  $\rho$  increases. Figure 3.6 shows that when the directionality  $\xi$  and the correlation coefficient  $\rho$  are increased in binomial networks, the critical threshold increases. The same behavior is observed in SF networks. In Fig. 3.5(a), we can also see that when the initial fraction  $f$  is close to 1, the normalized largest cluster  $s_1$  of opinion  $\sigma_+$  is close to 1, which means that the size of the largest cluster approximately equals to the size of the network. Hence, the algorithm ANC-DDR doesn't introduce evident disconnected components of in the generated networks. Note also that when the indegree outdegree correlation  $\rho = 1$ , the critical thresholds obtained

by DIR and ANC-DDR agree with each other for different values of the directionality  $\xi$ .

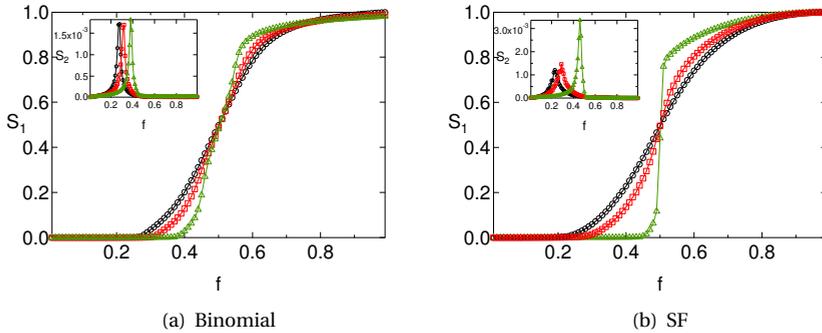


Figure 3.5: (Color online) Plot of the normalized largest cluster  $s_1$  of opinion  $\sigma_+$  as a function of the initial fraction  $f$ , when the directionality  $\xi = 0.6$ , for different values of the indegree and outdegree correlation  $\rho$ : 0( $\circ$ ), 0.5( $\square$ ), 1( $\triangle$ ), and for different networks with  $N = 10^5$  nodes and  $E[D] = 4$ : (a) Binomial, (b) SF. In the insets we plot  $s_2$  as a function of  $f$  with the same symbols and for the same networks as in the main figure. All results are based on averaging 1000 realizations.

The influence of the indegree and outdegree correlation  $\rho$  on the critical threshold can be understood as follows. A smaller  $\rho$  means a clearer inequality or asymmetry between the indegree and outdegree links. When the indegree and outdegree links are asymmetrical, a node with more in-neighbors than out-neighbors is more likely to follow the majority opinion and, because it has few out-neighbors, its own opinion will have little influence. Compared with the nodes which have the same number of in-neighbors and out-neighbors and tends to follow as well as spread the majority opinion, such nodes (with fewer out-neighbors) cannot help. Nodes with more out-neighbors than in-neighbors have greater influence and can thus hold the minority opinion and contribute to its spread. Thus the minority opinion benefits more from an inequality between the indegree and outdegree, or equivalently from a smaller  $\rho$ , so the lower correlation coefficient  $\rho$  leads to a smaller critical value  $f_c$ . Such imbalance between the number of friends that influence you and the number of friends that you can influence, actually, helps the existence of the minority opinion.

We now further explore the properties of nodes in the final steady state. We focus on binomial networks in the steady state and calculate as a function of  $\rho$ , the average indegree and outdegree in the largest  $\sigma_+$  and  $\sigma_-$  clusters with a directionality  $\xi = 1$  (generated by ANC) when the initial fraction  $f$  of opinion  $\sigma_+$  equals 0.4 (minority). As discussed above, and seen in Fig. 3.7, the outdegree links of a node with the minority opinion in the steady state tends to be larger for all  $\rho < 1$  than the indegree links, because nodes with few in-neighbors are less influenced by other nodes and thus can more easily keep their minority opinion. On the contrary, the indegree of a node with the majority opinion tends to be larger than its outdegree. Note that, when the initial fraction  $f$  of the opinion  $\sigma_+$  is 0.4, the average number of indegree links is smaller for the nodes in the largest  $\sigma_+$  cluster compared with the nodes in the largest  $\sigma_-$  cluster. Note also that in the majority clusters ( $\sigma^+$ ) both the indegree and the outdegree are close to 4, which is

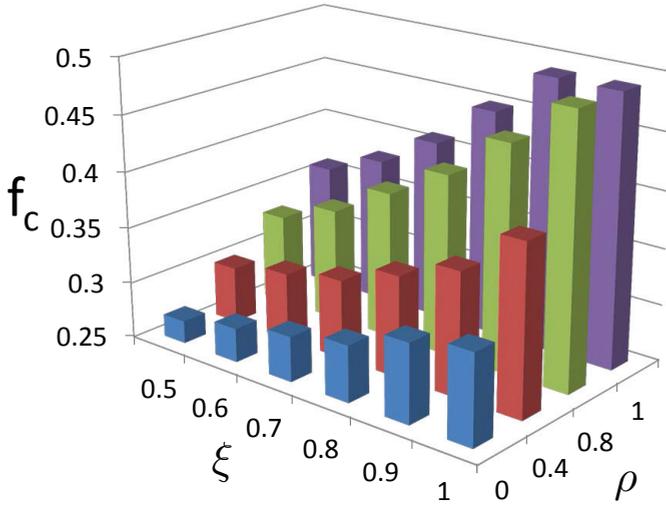


Figure 3.6: (Color online) Plot of the critical threshold  $f_c$  as a function of the linear correlation coefficient  $\rho$  and the directionality  $\xi$  for binomial networks. All results are based on averaging 1000 realizations.

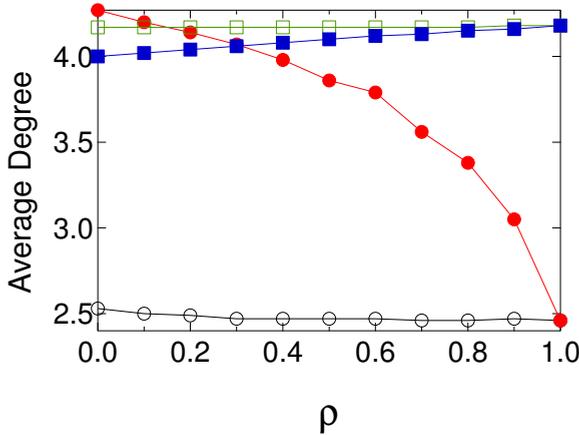


Figure 3.7: (Color online) Plot of the average indegree and outdegree of the nodes in the largest  $\sigma_+$  and  $\sigma_-$  cluster for binomial networks, when the initial fraction  $f$  of the opinion  $\sigma_+$  equals 0.4 as a function of  $\rho$ . The representation of the four lines are as follows: the average indegree (○) and outdegree (●) of the nodes in the largest  $\sigma_+$  cluster; the average indegree (□) and outdegree (■) of the nodes in the largest  $\sigma_-$  cluster. All results are based on averaging 100 realizations.

the average degree of the whole network. This is in marked contrast with the average indegree of the nodes in the largest minority cluster with degree approximately 2.5. The average outdegree of minority is larger than 4 when the linear correlation coefficient is  $\rho = 0$ . As  $\rho$  increases there is a higher correlation between the indegree and outdegree and the average outdegree of minority decreases rapidly.

### 3.6. CONCLUSIONS

Because of the ubiquity of the non-consensus steady state in real-world opinion competitions and the dominance of unidirectional relationships in real-world social networks, we study a non-consensus opinion model on directed networks. To quantify the extent to which a network is directed, we use a directionality parameter  $\xi$ , defined as the ratio between the number of unidirectional links and the total number of links. We also employ a linear correlation coefficient  $\rho$  between the indegree and outdegree to quantify any asymmetry.

We propose two approaches to construct directed networks. The first is directionality-increasing rewiring (DIR) and is used to rewire the links of an undirected network to obtain a directed network with any directionality value  $\xi$  without changing the indegree and outdegree, i.e., the indegree-outdegree correlation value is fixed at  $\rho = 1$ . The second is ANC-DDR, a combination of asymmetric indegree-outdegree network construction (ANC) and directionality-decreasing rewiring (DDR). Using ANC we construct a directed network ( $\xi \approx 1$ ) with an arbitrary but identical indegree and outdegree distribution and a given indegree-outdegree correlation  $\rho$ . We then use DDR to further decrease the directionality  $\xi$  of the network.

We use DIR and ANC-DDR to generate directed networks with a given combination of  $\xi$  and  $\rho$  and investigate how the directionality  $\xi$  and the linear correlation coefficient  $\rho$  between the indegree and outdegree links affect the critical threshold  $f_c$  of the NCO model. We find that in both binomial and SF networks increasing  $\xi$  or  $\rho$  increases the critical threshold  $f_c$ . We also find that as  $\xi$  and  $\rho$  increase, the phase transition becomes abrupt and is no longer second-order. We find that as a network becomes more directed it becomes more difficult for a minority opinion to form a cluster, while increasing the indegree-outdegree asymmetry makes the minority opinion more stable. Our work indicates that directionality and the asymmetry between indegree and outdegree play a critical role in real-world opinion competitions.



# 4

## EPIDEMIC SPREADING WITH DEGREE-CORRELATED HETEROGENEOUS RECOVERY RATES ON DIRECTED NETWORKS

*The nodes in communication networks are possibly and likely equipped with different recovery resources, which allow them to recover from a virus with different rates. In this chapter, we aim to know how to allocate the limited recovery resources to efficiently prevent the spreading of epidemics. Instead of considering the same recovery rate for all the nodes in the SIS model, we propose to allocate the recovery rate  $\delta_i$  for node  $i$  according to its indegree and outdegree- $\delta_i \sim k_{i,in}^{\alpha_{in}} k_{i,out}^{\alpha_{out}}$ , given the finite average recovery rate  $\langle \delta \rangle$  representing the limited recovery resources over the whole network. We consider directed networks with the same power law indegree and outdegree distribution, but different values of the directionality  $\xi$  and linear correlation coefficients  $\rho$  between the indegree and outdegree. We find that, by tuning the two scaling exponents  $\alpha_{in}$  and  $\alpha_{out}$ , we can always reduce the infection fraction  $y_\infty$  thus reducing the extent of infections, comparing to the homogeneous recovery rates allocation. Moreover, we can find our optimal strategy via the optimal choice of the exponent  $\alpha_{in}$  and  $\alpha_{out}$ . Our optimal strategy indicates that when the recovery resources are sufficient, more resources should be allocated to the nodes with a larger indegree or outdegree, but when the recovery resource is very limited, only the nodes with a larger outdegree should be equipped with more resources. We also find that our optimal strategy works better when the recovery resources are sufficient but not yet able to make the epidemic die out, and when the indegree outdegree correlation is small.*

---

This chapter have been published as: **B. Qu, A. Hanjalic, and H. Wang, Heterogeneous Recovery Rates against SIS Epidemics in Directed Networks, NetGCoop 2014: International Conference on NETWORK Games, CONTROL and OPTimization. Trento, Italy.**

## 4.1. INTRODUCTION

THE classic SIS model assumes that the infection rate  $\beta$  is the same for all infected-susceptible node pairs and so is the recovery rate  $\delta$ , thus homogeneity. However, the recovery time of real-world network components can be different, since they can be equipped with, for example, different antivirus softwares and different levels of human intervention that lead to different time to discover and cure the virus. Meanwhile, a fast recovery which requires an antivirus software with a high scanning frequency or frequent human intervention, is costly. The balance of the effect and cost of recovering capacity is of key importance. That is to say, with the limited recovery resources, how to distribute them, is critical.

In this chapter, we consider the SIS epidemic spreading with homogeneous infection rate  $\beta$  in a network with  $N$  nodes. However, we propose heterogeneous recovery rates allocation to minimize the fraction  $y_\infty$  of infected nodes in the steady state for the given finite average recovery rate  $\langle \delta \rangle = (\sum_{i=1}^N \delta_i) / N$  representing the limited recovery resources.

It has been revealed that in real-world networks such as the Internet[24] and World Wide Web[2], the probability that a node has  $k$  connections (degree  $k$ ) follows a scale-free distribution  $P(k) \sim k^{-\lambda}$ [29, 39, 113], with an exponent  $\lambda$  ranging between 2 and 3. The studies[10, 37] have also shown the directed characteristics of those networks: when a node  $i$  is connected to  $j$ ,  $j$  is not necessarily connected to  $i$ . Hence, we consider directed scale-free (SF) networks which has been less explored for the SIS model.

We propose to allocate the recovery rate of each node as a function of the indegree and outdegree of the node, i.e.  $\delta_i \sim k_{i,in}^{\alpha_{in}} k_{i,out}^{\alpha_{out}}$ . The degree, which is easy to compute, has been chosen for the following reasons: a) the nodal degree has been shown to be crucial to select e.g. the nodes to immunize to reduce the infection fraction and the nodes to protect to maintain the network connectivity[58, 119]; b) the degree is correlated with many other network properties of a node[68, 71]. When the network is undirected, our strategy becomes  $\delta_i \sim k_i^\alpha$ . Our goal is to find the optimal allocation, i.e. the optimal scaling parameter  $\alpha_{in}$  and  $\alpha_{out}$ , which leads to a minimal infection fraction. We find that, compared to the homogeneous SIS model, where the recovery rate is the same for all the nodes, our optimal strategy can always further evidently reduce the infection fraction  $y_\infty$  (even to 0, thus, the epidemic dies out) in the steady state. The novelty of our approach lies in: a. We generalize the classic homogeneous SIS model to heterogeneous, which allows each node to have a different recovery rate; b. We consider not only the classic undirected networks but also directed networks, characterized by the directionality  $\xi$  and the indegree outdegree correlation  $\rho$ . This introduces heterogeneity into the network structure; c. We propose a heterogeneous recovery rates allocation strategy based on the degree of each node, the simplest network property to compute.

The main method to reduce the fraction of the infected nodes is immunization. The immunization strategies[7, 31, 50, 82, 94] select a given number of nodes to be immunized. When a node gets immunized, it can not get infected nor infect the others. To immunize a node is equivalent to allocating an infinite recovery rate to this node. In this sense, the immunization problem is to decide which set of a given number  $m$  of nodes should we assign an infinite recovery rate whereas the rest nodes have the same given and finite recovery rate. Our work addressed the more general case where each node

may have a different recovery rate instead of assigning two possible recovery rates to the whole network. Heterogeneous SIS model has been addressed recently [23, 42, 96, 97] by taking into account either the heterogeneous infection or recovery rates. [42] studies the epidemic threshold with variable infection rates. [23] also models heterogeneous infection rates and observes slow epidemic extinction phenomenon. [96] and [97] discuss how to choose the infection and recovery rates from given discrete sets to let the virus die out, but our work addresses, more generally, how to reduce the fraction of infected nodes—a zero fraction means the extinction of the virus.

The rest of this chapter is organized as follows. Section 4.2 introduces the directionality  $\xi$  and indegree outdegree correlation  $\rho$  to describe the directed networks, and the epidemic threshold as well. Section 4.3 illustrates the distribution of the recovery rate  $\delta_i$ , the infection fraction vs. the epidemic threshold, and the influence of the exponents  $(\alpha_{in}, \alpha_{out})$  on the infection fraction, when the recovery rate is distributed heterogeneously according to our strategy. Section 4.4 investigates the optimal exponents  $(\alpha_{in,opt}, \alpha_{out,opt})$  for different indegree outdegree correlations  $\rho$ , and compares the infection fraction obtained by our optimal heterogeneous recovery rates allocation and by homogeneous recovery rates allocation. Section 4.5 concludes the paper.

## 4.2. DIRECTED NETWORKS AND THE EPIDEMIC THRESHOLD

### 4.2.1. DIRECTED NETWORKS

In order to describe the directed SF networks, as in chapter 3, we employ the directionality  $\xi$  and indegree outdegree correlation  $\rho$ . In this chapter, we confine ourselves to the case in which the indegree and outdegree follow the same distribution. In this case,  $\rho = 1$  implies that  $d_{i,in} = d_{i,out}$  holds for every node  $i$ . Particularly, in this chapter, we construct the directed SF networks with a given indegree outdegree correlation  $\rho$  using the algorithm ANC we proposed in chapter 3, so we get the networks with the directionality<sup>1</sup>  $\xi \approx 1$  and a specified indegree outdegree correlation  $\rho$ .

### 4.2.2. THE EPIDEMIC THRESHOLD

Given the adjacency matrix<sup>2</sup>  $A$  of a network, the epidemic threshold of the homogeneous SIS model with the same infection rate and recovery rate for all the links and nodes respectively, via the N-intertwined mean field approximation (NIMFA) is  $\tau_c = 1/\lambda_1(A)$ , where  $\lambda_1(A)$  is the largest eigenvalue of the adjacency matrix  $A$ [121]. When the effective infection rate  $\tau = \beta/\delta$  is above this threshold  $\tau > \tau_c$ , the epidemic spreads out. Via the mean field approximation-NIMFA, we can also obtain the sufficient condition for the epidemic to die out in the heterogeneous SIS model, which is  $\Re(\lambda_1(-diag(\delta_i) + \beta * A)) \leq 0$ [121][96], where  $A$  is the adjacency matrix of the network,  $diag(\delta_i)$  is a  $N$  by  $N$  diagonal matrix with elements  $\delta_i$  and  $\Re(\lambda_1(\bullet))$  represents the real part of the largest eigenvalue of a matrix<sup>3</sup>.

<sup>1</sup>The expectation of the directionality  $E[\xi] = 1 - E[D]^2 N / (N - 1)^2, \lim_{N \rightarrow \infty} E[\xi] = 1$ .

<sup>2</sup>The entries of the adjacency matrix  $A$  are,  $a_{ij} = 1$  if node  $i$  is connected to node  $j$  or otherwise  $a_{ij} = 0$ .

<sup>3</sup>If  $A$  is asymmetric, the largest eigenvalue of the matrix  $-diag(\delta_i) + \beta * A$  may be complex.

### 4.3. THE EFFECT OF THE HETEROGENEOUS RECOVERY RATES ALLOCATION

Instead of the approximated discrete-time SIS model simulations, we use a continuous-time simulator, proposed by van de Bovenkamp et al. and described in detail in [69]. We use the continuous-time simulations which precisely simulate the continuous SIS dynamics, instead of discrete-time ones, because the precision of a discrete time simulation of a continuous process requires a small time bin to sample the continuous process so that within each time bin, no multiple events occur. A heterogeneous SIS model allows different as well large infection or recovery rates, which requires even smaller time bin size and challenges the precision of discrete-time.

We are going to evaluate the effect of our degree based heterogeneous recovery rates allocation strategy via the infection fraction  $y_\infty$  in the steady state. In this chapter, we consider directed SF networks (both the indegree and outdegree follow the same power law distribution) with the power law exponent  $\lambda = 2.5$ , the networks size  $N = 1000$ , the smallest degree  $d_{\min} = 2$  and the natural degree cutoff  $d_{\max} = \lfloor N^{1/(\lambda-1)} \rfloor$  [30]. The directionality  $\xi$  and indegree outdegree correlation  $\rho$  can be tuned, using the algorithms we proposed in [99]. For the heterogeneous SIS model, we set the infection rate  $\beta = 1$  for all links, and assign the recovery rate according to our strategy  $\delta_i = c_2 d_{i,\text{in}}^{\alpha_{\text{in}}} d_{i,\text{out}}^{\alpha_{\text{out}}}$ , where the parameter  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$  can be tuned and the constant number  $c_2$  is decided by the given average recovery rate  $\langle \delta \rangle$ . All the results are based on the average of 1000 simulations.

#### 4.3.1. THE DISTRIBUTION OF THE RECOVERY RATE $\delta_i$

In this paper, the indegree and outdegree of a node both follow the same power-law distribution

$$\Pr[D_{\text{in}} = k] = \Pr[D_{\text{out}} = k] = c_1 k^{-\lambda}$$

where  $c_1 = 1 / \sum_{k=k_{\min}}^{k_{\max}} k^{-\lambda}$ . The joint degree distribution of a random node in a directed network with indegree outdegree correlation  $\rho$  and directionality  $\xi \approx 1$  that we generated using ANC follows

$$\Pr[D_{\text{in}} = k_{\text{in}}, D_{\text{out}} = k_{\text{out}}] = \begin{cases} \rho c_1 k_{\text{in}}^{-\lambda} + (1 - \rho) c_1^2 k_{\text{in}}^{-2\lambda}, & \text{when } k_{\text{in}} = k_{\text{out}} \\ (1 - \rho) c_1^2 k_{\text{in}}^{-\lambda} k_{\text{out}}^{-\lambda}, & \text{when } k_{\text{in}} \neq k_{\text{out}} \end{cases}$$

The recovery rate of each node is assigned according to our strategy as  $\delta_i = c_2 d_{i,\text{in}}^{\alpha_{\text{in}}} d_{i,\text{out}}^{\alpha_{\text{out}}}$ , where  $c_2 = \frac{\langle \delta \rangle}{\sum_{k_{\text{in}}=1}^{N-1} \sum_{k_{\text{out}}=1}^{N-1} \Pr[D_{\text{in}}=k_{\text{in}}, D_{\text{out}}=k_{\text{out}}] k_{\text{in}}^{\alpha_{\text{in}}} k_{\text{out}}^{\alpha_{\text{out}}}}$ , so the probability density function of the recovery rate  $\Delta$  of a random node is

$$\Pr[\Delta = x] = \sum_{k_{\text{in}}=1}^{N-1} \Pr[D_{\text{in}} = k_{\text{in}}, D_{\text{out}} = \left( \frac{x}{c_2 k_{\text{in}}^{\alpha_{\text{in}}}} \right)^{\frac{1}{\alpha_{\text{out}}}}].$$

In special cases, but not always, the distribution of the recovery rate follows strictly the power law. For example, when the indegree outdegree correlation is  $\rho = 1$ , the distribution follows the power law with the exponent  $\alpha_{\text{in}} + \alpha_{\text{out}} - \lambda$ . Fig. 4.1 shows an example of the distribution of the recovery rate with  $\langle \delta \rangle = 2$ ,  $\rho = 0.5$ ,  $\alpha_{\text{in}} = 0.3$  and  $\alpha_{\text{out}} = 0.6$ . The distribution in this case follows approximately a power law.

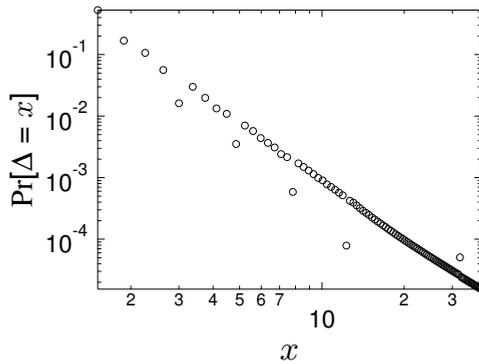


Figure 4.1: The probability distribution of the recovery rate  $\Delta$  of a random node with  $\langle \delta \rangle = 2$ ,  $\rho = 0.5$ ,  $\alpha_{in} = 0.3$  and  $\alpha_{out} = 0.6$ .

4.3.2. THE INFECTION FRACTION  $y_\infty$  VS.  $\Re(\lambda_1(-diag(\delta_i) + \beta * A))$

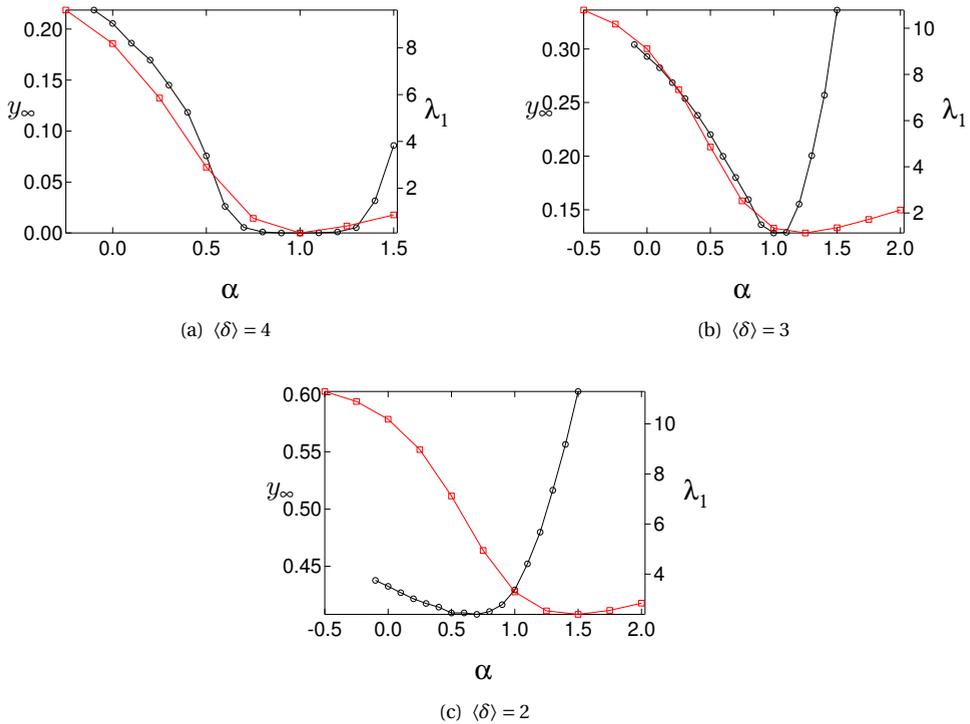


Figure 4.2: The plots of the infection fraction  $y_\infty$  (o) and the largest eigenvalue  $\lambda_1$  (□) of the matrix  $-diag(\delta_i) + A$  as a function of the exponent  $\alpha$  for different values of the average recovery rate  $\langle \delta \rangle$ .

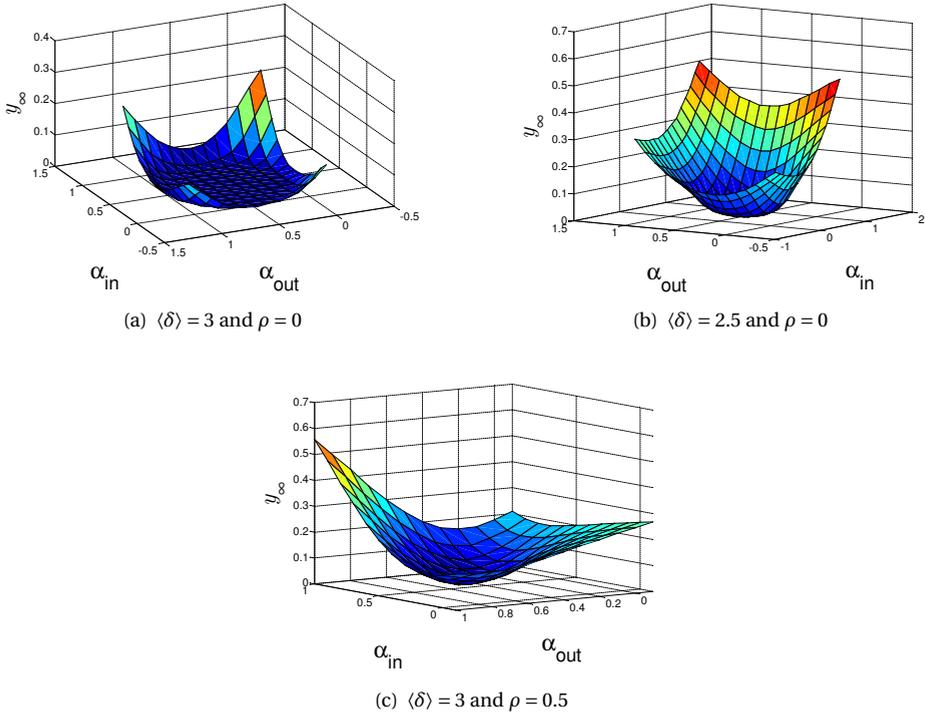


Figure 4.3: The plots of infected infraction  $y_\infty$  as a function of the exponents  $\alpha_{in}$  and  $\alpha_{out}$  with different values of the average recovery rate  $\langle \delta \rangle$  and the indegree outdegree correlation  $\rho$

From network designers' point of view, we aim to either increase the epidemic threshold or decrease the fraction of infection via network topology modifications, the allocation of the infection rates, recovery rates or immunization [70, 126, 128]. We would like to first understand in the heterogeneous SIS, whether a small  $\Re(\lambda_1(-diag(\delta_i) + \beta * A))$  would suggest a low infection fraction. If so, we could simplify the optimization of the fraction of infection to the optimization of the  $\Re(\lambda_1(-diag(\delta_i) + \beta * A))$ .

We take undirected scale-free networks as an example. In this case, the largest eigenvalue is real, i.e.  $\Re(\lambda_1(-diag(\delta_i) + \beta * A)) = \lambda_1(-diag(\delta_i) + \beta * A)$ . We compare the infection fraction  $y_\infty$  in the steady state and the largest eigenvalue  $\lambda_1$  of the matrix  $(-diag(\delta_i) + A)$ , when the infection rate  $\beta = 1$  holds for all the links, the recovery rate of each node is assigned according to our strategy, i.e.  $\delta_i = c_2 d_i^\alpha$  and the average recovery rate  $\langle \delta \rangle$  is given as shown in Fig. 4.2. In Fig. 4.2(a), we can see that when the infection fraction  $y_\infty = 0$ , the largest eigenvalue  $\lambda_1$  is also close to 0 which supports that  $\lambda_1(-diag(\delta_i) + \beta * A) < 0$  is the sufficient condition for an epidemic to die out. The trends of the two curves are consistent—the larger the eigenvalue  $\lambda_1$  is, the larger the infection fraction  $y_\infty$  is. However, their trends are not the same when the average recovery rate  $\langle \delta \rangle$  decreases (i.e. the recovery resources are reduced) as shown in Fig. 4.2(b) and

4.2(c). In other words, we cannot use the magnitude of  $\lambda_1$  to indicate the extent that the network is infected, which suggests that we have to design recovery allocation strategies to minimize the fraction of infection.

### 4.3.3. THE INFLUENCE OF $\alpha$ OR $(\alpha_{\text{in}}, \alpha_{\text{out}})$

We can also see in Fig. 4.2, compared to the homogeneous SIS model ( $\alpha = 0$ ), by applying our strategy to allocate the recovery rate, the infection fraction  $y_\infty$  can be significantly reduced in a wide range of  $\alpha$ , and we will discuss the maximum possible improvement in detail in the next section. Moreover, the infection fraction is again increasing if the exponent  $\alpha$  is larger than a certain value. That's to say, more recovery resources distributed to the high-degree nodes is not necessarily better in preventing infections; instead, too unbalanced distribution helps the virus. In the extreme case where  $\alpha \rightarrow +\infty$ , all the recovery rate is allocated to a single node, the one with the highest degree. In this case, all the other nodes have recovery rate 0. The network always gets fully infected except for the largest degree node, as long as the network is still connected after removing the highest degree node.

We investigate next the case of directed SF networks with directionality  $\xi \approx 1$  and a given indegree outdegree correlation  $\rho$ . Two scaling exponents  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$  can be tuned to decide the recovery rate  $\delta_i = c k_{i,\text{in}}^{\alpha_{\text{in}}} k_{i,\text{out}}^{\alpha_{\text{out}}}$ . We plot the infection fraction  $y_\infty$  as a function of  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$ . In Fig. 4.3(a), compared to the undirected networks with the same average recovery rate  $\langle \delta \rangle = 3$ , we find that most combinations of the two exponents  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$  around  $(0.5, 0.5)$ , i.e.  $\alpha_{\text{in}} \approx 0.5, \alpha_{\text{out}} \approx 0.5$ , can lead to the die-out steady state. If we further decrease the average recovery rate  $\langle \delta \rangle$ , though the die-out steady state can not be reached any more, many combinations of  $(\alpha_{\text{in}}, \alpha_{\text{out}})$  can reduce the infection fraction effectively compared to the homogeneous case, where  $\alpha_{\text{in}} = \alpha_{\text{out}} = 0$ , as shown in Fig. 4.3(b). Fig. 4.3(c) describes the case with a smaller average recovery rate  $\langle \delta \rangle$  but larger indegree outdegree correlation ( $\rho = 0.5$ ), the same phenomenon can be observed.

## 4.4. OPTIMAL HETEROGENEOUS RECOVERY ALLOCATIONS

Most interestingly, we would like to find out our optimal strategy, i.e. the optimal exponent  $\alpha_{\text{opt}}$  (or a combination  $(\alpha_{\text{in,opt}}, \alpha_{\text{out,opt}})$ ) which leads to the minimum infection fraction  $y_\infty$ . Moreover, how much can our optimal strategy further reduce the fraction  $y_\infty$  of infection compared to the homogeneous allocation of recovery rate?

### 4.4.1. THE OPTIMAL EXPONENTS $\alpha$ OR $(\alpha_{\text{in}}, \alpha_{\text{out}})$

We list the optimal exponent  $\alpha$  for undirected SF networks with different average recovery rates  $\langle \delta \rangle$  in Table 4.1, and the optimal combination  $(\alpha_{\text{in,opt}}, \alpha_{\text{out,opt}})$  of the two exponents  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$  for directed SF networks with different average recovery rates  $\langle \delta \rangle$  and indegree outdegree correlation  $\rho$  in Table 4.2. All the optimal exponents are obtained by simulations, i.e. given  $\langle \delta \rangle$  and  $\rho$ , we run the simulations with different values of  $\alpha$  and select the exponent  $\alpha$  which leads to the minimum infection fraction  $y_\infty$  as the optimal exponent. For some combinations of  $\langle \delta \rangle$  and  $\rho$ , like  $\langle \delta \rangle = 3$  and  $\rho = 0$ , more than one pair of  $(\alpha_{\text{in}}, \alpha_{\text{out}})$  can reduce the infection fraction  $y_\infty$  to 0, so we cannot list the optimal one. Note that, when indegree outdegree correlation  $\rho = 1$ , the indegree and

outdegree are the same for each node, so it is enough to list only one optimal exponent. We still confine ourselves to directed networks with directionality  $\xi \approx 1$ .

Table 4.1: The optimal exponent  $\alpha_{opt}$  for undirected SF networks with different average recovery rates  $\langle\delta\rangle$

$\langle\delta\rangle$	1.5	1.75	2	2.25	2.5	2.75~4
$\alpha_{opt}$	0.3	0.4	0.7	0.8	0.9	1

Table 4.2: The optimal combinations  $(\alpha_{in,opt}, \alpha_{out,opt})$  for directed SF networks with different  $\langle\delta\rangle$  and  $\rho$

$\langle\delta\rangle \backslash \rho$	0	0.5	1
1.5	(0.1, 0.6)	(-0.1, 0.6)	0.1
1.75	(0.2, 0.6)	(0.1, 0.6)	0.3
2	(0.4, 0.5)	(0.3, 0.6)	0.5
2.25	(0.5, 0.5)	(0.4, 0.5)	0.7
2.5	(0.5, 0.5)	(0.5, 0.5)	0.8
2.75	-	(0.5, 0.5)	0.9
3	-	-	1

By observing Table 4.1 and 4.2, we find that, with indegree outdegree correlation  $\rho = 1$ , either the network is undirected or directed, the optimal exponent  $\alpha_{opt}$  decreases as the average recovery rate  $\langle\delta\rangle$  decreases. A positive  $\alpha_{opt}$  implies that, we should distribute more recovery resources to the nodes with larger degrees. As the total recovery resources,  $\langle\delta\rangle$  decrease,  $\alpha_{opt}$  decreases, suggesting a relative homogeneous allocation of the recovery resources. When the recovery resources are large,  $\alpha_{opt}$  is large, implying that a more heterogeneous distribution is most beneficial. Moreover, for directed networks with unequal indegree and outdegree, regardless of the average recovery rate  $\langle\delta\rangle$ , the optimal exponent  $\alpha_{out,opt}$  for outdegree is always around 0.5, though the optimal exponent  $\alpha_{in,opt}$  for indegree declines as  $\langle\delta\rangle$  decreases. In other words, the nodes with larger outdegree should always be equipped with faster recovery even if the total recovery resources are very limited, but not the nodes with larger indegree. That's reasonable because a larger outdegree means more chances to spread the virus and a fast recovery of such nodes is effective to prevent infections.

#### 4.4.2. OPTIMAL HETEROGENEOUS VS. HOMOGENEOUS RECOVERY RATES

Furthermore, we would like to understand how much our optimal strategy could further reduce the infection compared with classic homogeneous recovery rates allocation. We compare the infection fraction  $y_\infty$  of the homogeneous ( $\alpha = 0$  or  $\alpha_{in} = \alpha_{out} = 0$ ) and our optimal heterogeneous ( $\alpha = \alpha_{opt}$  or  $\alpha_{in} = \alpha_{in,opt}, \alpha_{out} = \alpha_{out,opt}$ ) recovery rates allocation.

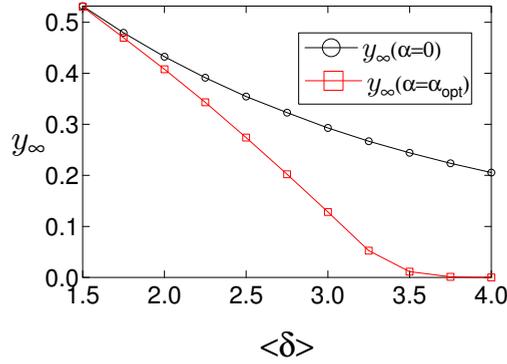


Figure 4.4: The plot of the infection fraction  $y_\infty$  as a function of the average recovery rate  $\langle \delta \rangle$  for undirected SF networks under both homogeneous ( $\alpha = 0$ ) and optimal ( $\alpha = \alpha_{opt}$ ) recovery rates allocation.

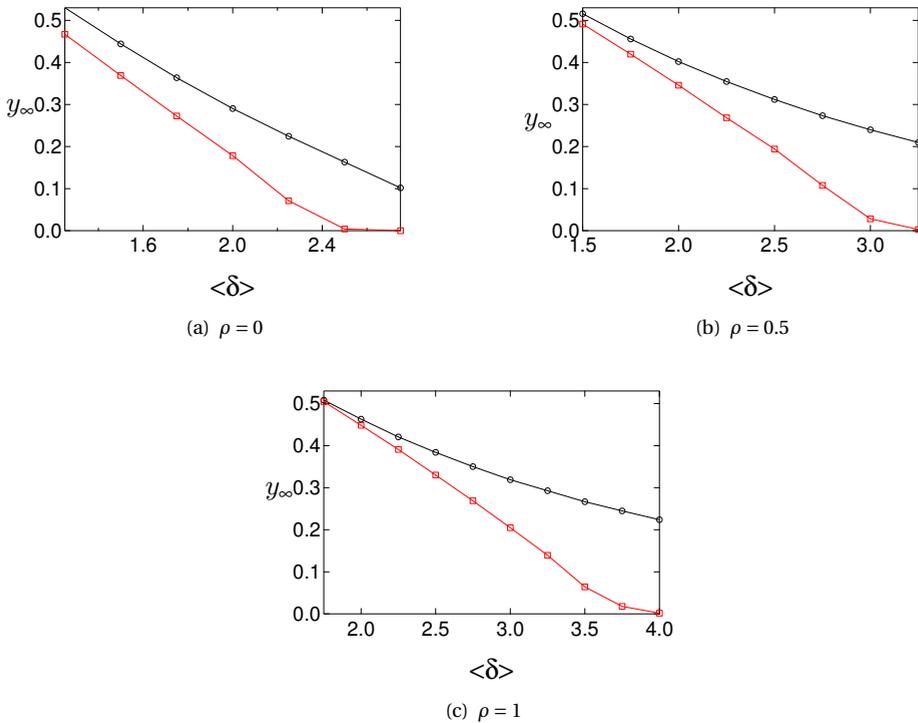


Figure 4.5: The infection fraction  $y_\infty$  as a function of the average recovery rate  $\langle \delta \rangle$  in directed SF networks with different indegree outdegree correlation  $\rho$ , under both homogeneous ( $\circ$ ,  $\alpha_{in} = \alpha_{out} = 0$ ) and optimal ( $\square$ ,  $\alpha_{in} = \alpha_{in, opt}$ ,  $\alpha_{out} = \alpha_{out, opt}$ ) recovery rates allocation.

Fig. 4.4 shows the results for undirected SF networks, and Fig. 4.5 for the directed SF networks with different indegree outdegree correlation  $\rho$ . We find that, in general, our optimal heterogeneous strategy outperforms the homogeneous one and such out-performance becomes more evident when the average recovery rate  $\langle\delta\rangle$  is not too small (not close to the infection rate  $\beta = 1$ ). Our optimal heterogeneous strategy could even reduce the outbreak epidemic state with the homogeneous strategy into a die out state when the average recovery rate  $\langle\delta\rangle$  is close to 3.

Fig. 4.5 shows that a smaller indegree outdegree correlation  $\rho$  retards the virus spreading given the recovery rate allocation strategy, e.g. the homogeneous allocation. A smaller indegree outdegree correlation  $\rho$  implies the larger difference between the indegree and outdegree of nodes. On one hand, a node with large outdegree may have a small indegree. Though such nodes have more chances to spread the virus, they are not likely to be infected. On the other hand, a node with a small outdegree may have a large indegree. Such nodes tend to be infected but do not help the virus to spread. Hence, a smaller indegree outdegree correlation  $\rho$  prevents the spreading of the virus.

We further explore the performance of our optimal heterogeneous recovery rates allocation strategy for different indegree outdegree correlation  $\rho$ . We calculate the further improvement in the infection fraction,  $\Delta y_\infty$ , as the difference between the infection fraction of the homogeneous ( $y_{\infty,\text{homo}}$ ) and our optimal heterogeneous ( $y_{\infty,\text{opt}}$ ) recovery rates allocation, i.e.  $\Delta y_\infty = y_{\infty,\text{homo}} - y_{\infty,\text{opt}}$ , given the same average recovery rate  $\langle\delta\rangle$ . In Fig. 4.6, for different indegree outdegree correlation  $\rho$ , we plot  $\Delta y_\infty$  as a function of the average recovery rate  $\langle\delta\rangle$  to show how much our optimal strategy can reduce the infection fraction, compared to the homogeneous SIS model. There seems to be a peak at  $\langle\delta\rangle_c$  in each curve. That's mainly because as  $\langle\delta\rangle$  increases, both  $y_{\infty,\text{homo}}$  and  $y_{\infty,\text{opt}}$  decrease to 0, limiting the difference between the heterogeneous and homogeneous strategy. Before the peak, i.e. when  $\delta < \langle\delta\rangle_c$ , we find that the effect of our optimal strategy is more evident when the average recovery rate  $\langle\delta\rangle$  is large and when the indegree outdegree correlation  $\rho$  is small. When the total recovery resources, i.e. the average recovery rate is very small or very large, the epidemic will anyway spread out or die out respectively, which is independent of the recovery rate allocation strategy. Whereas in between these two extreme cases, i.e. with intermediate  $\langle\delta\rangle$ , we always see the outperformance of our heterogeneous strategy. Such outperformance appears in a range of  $\langle\delta\rangle$  with smaller values when the indegree outdegree correlation  $\rho$  is smaller, likely due to the fact that a smaller degree correlation contributes to the prevention of epidemic spreading.

## 4.5. CONCLUSION

In this chapter, we address a new challenging question: how to allocate the limited recovery resources heterogeneously so that the fraction of infection can be minimized? We propose the heterogeneous recovery rates allocation strategy which allocates different recovery rates to different nodes. Our strategy is based on the degree of each node, which has the lowest computational complexity,  $\delta_i = ck_{i,\text{in}}^{\alpha_{\text{in}}} k_{i,\text{out}}^{\alpha_{\text{out}}}$ . We consider both undirected and directed networks, characterized by the directionality  $\xi$  and the indegree outdegree correlation  $\rho$ .

Interestingly, our strategy via the optimal choice of the parameters  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$ , evidently outperforms the classic homogeneous allocation of recovery resources in gen-

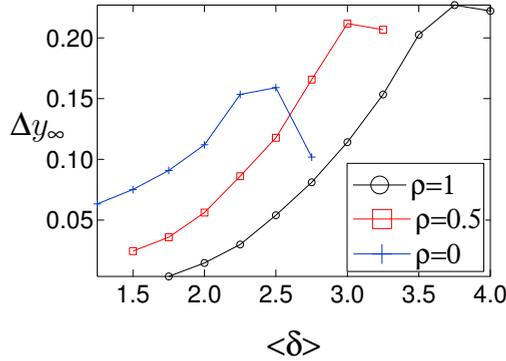


Figure 4.6: The variation of the infection fraction  $\Delta y_\infty$  as a function of the average recovery rate  $\langle \delta \rangle$  for different values of the indegree outdegree correlation  $\rho$ .

eral, especially when the given recovery resources are sufficient. The optimal choice of the parameter  $\alpha_{\text{in}}$  and  $\alpha_{\text{out}}$  depends on the indegree outdegree correlation  $\rho$  and average recovery rate  $\langle \delta \rangle$ . We find that in undirected networks, when the average recovery rate  $\langle \delta \rangle$  is large,  $\alpha_{\text{opt}}$  is large, meaning that we should allocate the the recovery resources more heterogeneously. In directed networks, it seems that we should allocate more recovery resources to the high outdegree nodes, whereas the indegree has minor influence in determining the recovery rate as the average recovery rate  $\langle \delta \rangle$  decreases.

Our degree based heterogeneous recovery rates allocation strategy illustrates the potential to more effectively reduce infection than the classic homogeneous allocation. However, this is just a start and hopefully it could inspire better heterogeneous strategies. It would be interesting to further consider the heterogeneous infection rate, motivated by real-world data set. The allocation of recovery rates in that case is far from well understood.



# 5

## EPIDEMIC SPREADING WITH I.I.D. HETEROGENEOUS INFECTION RATES

*In this chapter, we aim to understand the influence of the heterogeneity of infection rates on the Susceptible-Infected-Susceptible (SIS) epidemic spreading. Employing the classic SIS model as the benchmark, we study the influence of the independently identically distributed infection rates on the average fraction of infected nodes in the metastable state. The log-normal, gamma and a newly designed distributions are considered for infection rates. We find that, when the recovery rate is small, i.e. the epidemic spreads out in both homogeneous and heterogeneous cases: 1) the heterogeneity of infection rates on average retards the virus spreading, and 2) a larger even-order moment of the infection rates leads to a smaller average fraction of infected nodes, but the odd-order moments contribute in the opposite way; when the recovery rate is large, i.e. the epidemic may die out or infect a small fraction of the population, the heterogeneity of infection rates may enhance the probability that the epidemic spreads out. Finally, we verify our conclusions via real-world networks with their heterogeneous infection rates. Our results suggest that, in reality the epidemic spread may not be so severe as the classic SIS model indicates, but to eliminate the epidemic is probably more difficult.*

---

This chapter have been published as: **B. Qu and H. Wang, SIS Epidemic Spreading with Heterogeneous Infection Rates, IEEE Transactions on Network Science and Engineering, Issue 99, 2017.**

## 5.1. INTRODUCTION

THE studies on contagion processes in networks are strongly motivated and justified by the anticipated outbreaks of epidemic diseases in a population and non-stop threats of cyber security in computer networks [3, 45, 46, 60, 83]. The Susceptible-infected-susceptible (SIS) model [15, 28, 33, 69, 111, 121, 128] is one of the most widely used models to describe such processes. In the continuous-time Markovian SIS model, a node is either infected or susceptible at any time  $t$ . Each infected node infects each of its susceptible neighbors with an infection rate  $\beta$ . The infected node can be recovered with a recovery rate  $\delta$ . Both infection and recovery processes are independent Poisson processes. The average fraction  $y_\infty$  of the infected nodes in the metastable state, ranging in  $[0, 1]$ , indicates how severe the influence of the virus is: the larger  $y_\infty$  is, the more severely the network is infected.

The classic SIS model assumes that the infection rate  $\beta$  is the same for all infected-susceptible node pairs and so is the recovery rate  $\delta$  for all nodes. Most studies are focusing on the relationship between the effective infection rate  $\tau$  and the average fraction  $y_\infty$  of infected nodes or the epidemic threshold in the virus contamination process with homogeneous infection (recovery) rates. However, in reality, neither the contact frequency [40] between a pair of individuals in social networks nor the connecting frequency between a pair of nodes in computer networks is constant. Infection rates can be different from pairs to pairs, thus heterogeneous. Many studies on real diseases, such as SARS [129] and Plasmodium falciparum infection [114] also reveal the heterogeneity of infection rates. Furthermore, Smith et al. [114] suggest that the distribution of infection rates in different populations may be varied as well, and Wang et al. [129] find that infection rates with the log-normal distribution fit best the data of SARS in 2003 by applying their model.

In this paper, we explore the effect of heterogeneous infection rates on the average fraction  $y_\infty$  of infected nodes in a systematic way. We propose a SIS model, in a network with  $N$  nodes, with the homogeneous recovery rate  $\delta$  but heterogeneous infection rates  $\beta_{ij}$  ( $= \beta_{ji}$ ,  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, N$  and  $i \neq j$ ) between node  $i$  and node  $j$ . Similar to the classic homogeneous SIS, our SIS model with heterogeneous infection rates is as well a Markovian process where the time for an infected node  $i$  to infect each of its susceptible neighbors  $j$  is an independent exponential random variable with average  $\beta_{ij}^{-1}$ . The homogeneous SIS model has the same infection rate  $\beta$  for all node pairs whereas all the infection rates in our heterogeneous SIS are independent and identically distributed (i.i.d.) random variables. We study how the distribution of infection rates influences the average fraction  $y_\infty$  of infected nodes in the metastable state.

A few recent papers [23, 42, 96–98, 135] have taken into account either the heterogeneous infection or recovery rates. In [98], we explored the influence of degree-based recovery rates on the average fraction of infected nodes in the metastable state. Preciado et al. [96, 97] discussed how to choose the infection and recovery rates from given discrete sets to let the virus die out. Fu et al. [42] studied the epidemic threshold when the infection rates depend on the node degrees and Buono et al. [23] considered a specific distribution of infection rates and observed slow epidemic extinction phenomenon. Yang and Zhou [135] gave an edge-based mean-field solution of the epidemic threshold in regular networks (the degrees of all nodes are the same) with i.i.d. heterogeneous in-

fection rates (following uniform or power-law distribution).

In this paper, we explore the influence of heterogeneous infection rates on the epidemic spreading. In practice, the number of new infections in a period of time can be used to estimate the infection rate, for example, [103] counts the number of infected people per time interval (daily, weekly, etc.) to indicate the infection rate; [32], illustrating a strategy to estimate the time-varying transmission rates for the spread of infection, also takes into account the daily distribution of new infections. Besides the number of new infections, the interacting frequencies between two neighboring nodes have also been employed to estimate the infection rate, for example, the infection rate has been considered to be proportional to the interacting frequency. The average infection rate obtained in both scenarios has been used as the infection rate in the homogeneous epidemic model. Our work points out how such assumption of homogeneous rates would differ from real-world heterogeneous infection rates with respect to their influence on the fraction of infected population. We consider several representative distributions with the same mean but higher moments tunable, since the influence of the mean<sup>1</sup> has been widely studied in the homogeneous SIS model [53, 72, 91, 92, 95]. To our best knowledge, our work is the first to discuss the influence of higher moments of the infection rate distribution in epidemic models.

## 5.2. SIS MODEL WITH HETEROGENEOUS INFECTION RATES

In this section, we introduce the classic SIS model, basic network models, the heterogeneous infection rates and the simulation settings of the SIS model with heterogeneous infection rates on a network.

### 5.2.1. NETWORK CONSTRUCTION

In real-world networks, the exponent  $\lambda$  is usually in the range [2,3], thus we confine the exponent  $\lambda = 2.5$  in this chapter. We further employ the smallest degree  $d_{\min} = 2$ , the natural degree cutoff  $d_{\max} = \lfloor N^{1/(\lambda-1)} \rfloor$  [30], and the size  $N = 10^4$ . Hence, the average degree is approximately 4. As the comparison, we consider the ER networks with the size  $N = 10^4$  and the average degree  $E[D] = 4$ .

### 5.2.2. INDEPENDENT AND IDENTICALLY DISTRIBUTED HETEROGENEOUS INFECTION RATES

In this subsection, we introduce three distributions of the heterogeneous infection rates. We aim to explore how the heterogeneous infection rates influence the spread of SIS epidemics, particularly we study the relationship between the variance<sup>2</sup> (and even higher moments) of the heterogeneous infection rates and the average fraction  $y_\infty$  of infected nodes. Hence, we would like to choose infection-rate distributions systematically such that they cover a broad range of distributions including those observed in real-world and importantly their higher order moments, at least the variances are tunable when their means are fixed.

The  $n$ th moment  $m_n$  of a distribution with the probability density function (PDF)

<sup>1</sup>The infection rate between any pair of nodes equals to the mean in the homogeneous SIS model.

<sup>2</sup>The variance of a random variable is the second central moment.

$f_B(\beta)$  is  $m_n = \int_{-\infty}^{+\infty} \beta^n f_B(\beta) d\beta$ . Thus, the first moment  $m_1$  is just the mean and the relationship between the second moment  $m_2$  and variance  $Var[B]$  is  $Var[B] = m_2 - m_1^2$ , where the random variable  $B$  is the infection rate of a link. To eliminate the influence of the mean  $m_1$ , we further define the  $n$ th normalized moment  $v_n = \frac{m_n}{m_1^n}$ , then  $v_1 = 1$  and the normalized variance  $v = v_2 - 1$ .

We choose two asymmetric distributions: the log-normal and gamma distribution, of which we can keep the means unchanged and tune the variances in a large range. The log-normal distribution [123]  $B \sim \text{Log-}\mathcal{N}(\beta; \mu, \sigma)$ , of which the PDF is, for  $\beta > 0$

$$f_B(\beta; \mu, \sigma) = \frac{1}{\beta \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln \beta - \mu)^2}{(2\sigma^2)}\right)$$

and the  $n$ th normalized moment is  $v_n = \exp(\frac{(n^2-n)\sigma^2}{2})$ , has a power-law tail for a large range of  $\beta$  provided  $\sigma$  is sufficiently large. The log-normal distribution has as well been widely observed in real-world, where the interaction frequency between nodes is usually considered as the infection rate between those nodes. One example is the infection rates of the co-author network, as illustrated in Fig. 5.8(b), Section 5. Moreover, Wang et al. [129] find that by employing the log-normal distributed infection rates, their epidemic model can accurately fit the infection data of 2003 SARS.

The gamma distribution  $B \sim \Gamma(\beta; k, \theta)$ , of which the PDF is, for  $\beta > 0$

$$f_B(\beta; k, \theta) = \exp\left(-\frac{\beta}{\theta}\right) \frac{\beta^{k-1}}{\theta^k \Gamma(k)}$$

( $\Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt$ ) and the  $n$ th normalized moment is  $\prod_{i=0}^{n-1} (1 + i k^{-1})$ , has a lighter tail than the log-normal distribution. The Airline network, as demonstrated in Fig. 5.8(a), has an exponentially distributed infection rates, which corresponds to the Gamma distribution when  $k = 1$ .

In order to take into account symmetrically distributed infection rates as well, we design a variance-tunable and symmetric distribution other than the two asymmetric distributions above. We call it the symmetric polynomial (SP) distribution  $B \sim \text{SP}(\beta; a, b)$ , whose PDF is

$$f_B(\beta; a, b) = \frac{b(a+1)}{2} |\beta - 1|^a$$

where  $\beta \in [1 - \frac{1}{\sqrt{b}}, 1 + \frac{1}{\sqrt{b}}]$  and,  $a = 1$  and  $b \in [1, +\infty)$  or  $b = 1$  and  $a \in [1, +\infty)$ . The mean of the distribution is 1, the variance is  $\frac{a+1}{b(a+3)}$ . Compared to the commonly-used uniform distribution (also symmetric and variance-tunable) with the same mean, the SP distribution can be tuned in a larger range of the variance.

### 5.2.3. THE SIMULATIONS

In order to study the effect of the variance of the heterogeneous infection rates on the virus spread, we perform simulations to obtain the fraction  $y_{\infty}$  of infected nodes as a function of the normalized variance  $v$  of infection rates on both ER and SF networks.

We find that, for commonly used 2-parameter distributions (such as the uniform distribution, log-normal distribution, gamma distribution, etc.), the scaling on the mean of infection rates can be eliminated by the same scaling on the recovery rate if we keep the normalized variance  $\nu$  unchanged. This conclusion is also consistent with the fact that only the effective infection rate  $\frac{\beta}{\delta}$  matters for the epidemic spreading, but not the infection rate  $\beta$  in the homogeneous SIS model. Hence, without loss of generality, we set the mean  $m_1$  of the infection rates to 1, thus all the normalized moments  $\nu_n$  equal to the unnormalized ones  $m_n$ .

### 5.3. SMALL RECOVERY RATES

In this work, the average of the heterogeneous infection rates and the homogeneous infect rate are the same. Since the recovery rate  $\delta$  plays the key role in the epidemic spreading, we discuss our results according to different ranges of the recovery rates. In this section, we introduce our main results about how the heterogeneous infection rates influence the contagion processes of epidemic, when the recovery rates are small such that the epidemic spreads out in both homogeneous and heterogeneous cases. In the next section, we focus on large recovery rates – the homogeneous effective infection rate  $\tau$  is close to the epidemic threshold  $\tau_c$  in the classic model, where the epidemic with homogeneous infection rates may die out.

#### 5.3.1. THE OBSERVATIONS

We first show the simulation results when the variance of the infection rates is smaller than 1, since the variance of a non-negative and symmetric distribution cannot be larger than the square of its mean<sup>3</sup>, thus 1 in this paper.

In Fig. 5.1, we find that the average fraction  $y_\infty$  of infected nodes decreases as the variance  $\nu$  of the infection rates increases, no matter which distribution the infection rates follow. Moreover, the comparison of the decay of the three curves in Fig. 5.1 also suggests that, the smaller the third moment<sup>4</sup> of the infection rate distribution is, the faster  $y_\infty$  decays as the variance increases.

When the variance  $\nu$  is larger than 1, the infection rates cannot be symmetrically distributed. We thus discuss only the log-normal and gamma distributions which are representative among the heavy-tailed distributions and widely used in the real-world analysis.

In Fig. 5.2, we observe the same as in Fig. 5.1. Moreover, we find that the average fraction  $y_\infty$  of infected nodes decays much faster when infection rates follow gamma distributions than log-normal distributions.

Here we only show the simulation results on ER random networks with  $10^4$  nodes and average degree  $E[D] = 4$ , because simulation results on SF networks lead to the same observations. Moreover, though not shown in this paper, we have also done the simulations

<sup>3</sup>For any random variable  $B$  following a non-negative and symmetric distribution  $f_B(\beta)$  with mean  $m_1$ , the smallest and largest value that  $B$  can reach is 0 and  $2m_1$  respectively, so the largest variance, which equals to  $m_1^2$ , can be reached when  $Pr[B = 0] = Pr[B = 2m_1] = 0.5$ .

<sup>4</sup>The third moment of the log-normal, gamma and SP distribution is  $(\nu + 1)^3$ ,  $(\nu + 1)(2\nu + 1)$  and  $3\nu + 1$  respectively.

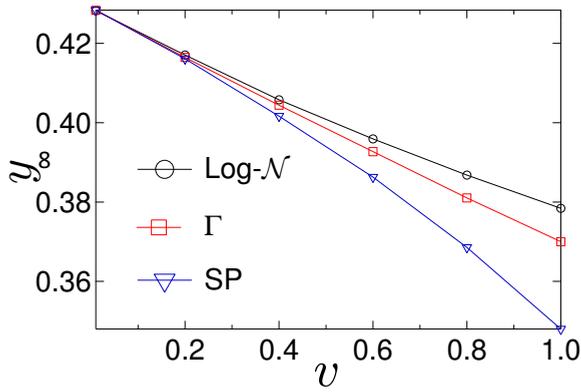


Figure 5.1: The average fraction  $y_\infty$  of infected nodes as a function of the normalized variance  $\nu$  of infection rates for log-normal ( $\circ$ ), gamma ( $\square$ ), and SP ( $\nabla$ ) infection-rate distributions respectively, and the recovery rate  $\delta = 2$ . We consider ER networks with average degree  $E[D] = 4$  and network size  $N = 10^4$ . The results are averaged over 1000 realizations.

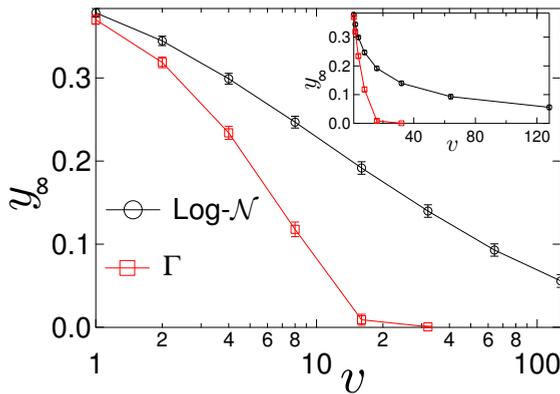


Figure 5.2: The average fraction  $y_\infty$  of infected nodes as a function of the variance  $\nu$  of infection rates following different distributions: log-normal ( $\circ$ ) and gamma ( $\square$ ), and the recovery rate  $\delta = 2$ . The simulations are on ER networks with average degree  $E[D] = 4$  and network size  $N = 10^4$ . The results are averaged over 1000 realizations, and the error bars are the standard deviations of the results in different realizations. The inset is the same as the main graph, but in a linear-linear scale.

with various values of the recovery rate, such as  $\delta = 0.1, 0.2, 1$ , etc., for both ER and SF networks and the conclusions are consistent.

### 5.3.2. THE INFLUENCE OF THE MOMENTS OF THE INFECTION RATES

To explain our observations, we consider a susceptible node and an infected node interconnected by a link. The probability  $\rho(T)$  that the infected node infects the susceptible neighbor in an arbitrary period  $T$ , is  $\rho(T) = \int_0^{+\infty} f_B(\beta)F(T; \beta) d\beta$ , where  $f_B(\beta)$  is the PDF of the infection rate, and  $F(T; \beta)$  is the probability that infection occurs between the neighboring infected and susceptible node pair within the time interval  $T$  when the infection rate is  $\beta$ . Since the infection between any infected and susceptible node pair is an independent Poisson process, the time for an infected node to infect a susceptible neighbor is an exponential variable, *i.e.*  $F(T; \beta) = 1 - e^{-T\beta}$ . We consider further the classic homogeneous SIS model, whose infection rate is equal to the average infection rate  $E[B]$  in our heterogeneous SIS model. The counterpart of  $\rho(T)$  in the homogeneous SIS model is, then,  $\rho^*(T) = F(T; E[B])$ .

**Theorem 1** *If  $f_B(\beta)$  is the probability density function of a non-negative continuous random variable  $B$ , and  $F(T; \beta)$  is the distribution function of an exponential random variable with the rate parameter  $\beta$ , then for any  $T > 0$ , we have*

$$\int_0^{\infty} f_B(\beta)F(T; \beta) d\beta \leq F(T; E[B])$$

**Proof.**

$$\begin{aligned} & \int_0^{\infty} f_B(\beta)F(T; \beta) d\beta \\ &= 1 - \int_0^{\infty} f_B(\beta)e^{-T\beta} d\beta \\ &= 1 - E[e^{-TB}] \end{aligned}$$

Since the exponential function is convex, Jensen's inequality [123] tells us that

$$E[e^{-TB}] \geq e^{-TE[B]}$$

Hence,

$$\int_0^{\infty} f_B(\beta)F(T; \beta) d\beta \leq 1 - e^{-TE[B]} = F(T; E[B])$$

■

Theorem 1, that proves  $\rho(T) \leq \rho^*(T)$ , tells us that if the infection rate in the classic homogeneous SIS model and the average infection rate in heterogeneous model are the same, then in the same period of time an infection event is more likely to happen in the classic SIS model.

We define  $\chi(T) = \rho^*(T) - \rho(T)$  as the difference in infection probability within an arbitrary time interval  $T$  between the SIS model with homogeneous and heterogeneous

infection rates.

$$\begin{aligned}
 \chi(T) &= E[e^{-TB}] - e^{-TE[B]} \\
 &= \sum_{n=0}^{\infty} \frac{(-T)^n (E[B^n] - (E[B])^n)}{n!} \\
 &= \sum_{n=0}^{\infty} \frac{(m_n - m_1^n)(-T)^n}{n!} \\
 &= \sum_{n=0}^{\infty} (v_{2n} - 1) \frac{(Tm_1)^{2n}}{(2n)!} - \sum_{n=0}^{\infty} (v_{2n+1} - 1) \frac{(Tm_1)^{2n+1}}{(2n+1)!}
 \end{aligned} \tag{5.1}$$

Note that the first step in (5.1) is valid only if the sum  $\sum_{n=0}^{\infty} \frac{(-T)^n E[B^n]}{n!}$  converges. The general log-normal distribution over an infinite range does not satisfy this condition. However, the infection rates of real-world systems are finite. Theorem 2 states that any realistic distribution of the infection rates within a finite range satisfies this convergence condition.

5

**Theorem 2** For any non-negative random variable  $B$  distributed in a finite range  $[0, b]$  and any finite  $T$ , the sum

$$\sum_{n=0}^{\infty} \frac{(-T)^n E[B^n]}{n!} \leq 2e^{Tb}$$

thus converges.

**Proof.**

$$\begin{aligned}
 E[B^n] &= \int_0^b \beta^n f_B(\beta) d\beta \\
 &= \beta^n \int_0^\beta f_B(\beta) d\beta \Big|_0^b - \int_0^b \int_0^\beta f_B(\beta) d\beta d\beta^n \\
 &= \beta^n F_B(\beta) \Big|_0^b - \int_0^b F_B(\beta) d\beta^n
 \end{aligned}$$

Since

$$F_B(\beta) = \int_0^\beta f_B(\beta) d\beta \leq 1$$

we have

$$E[B^n] \leq b^n + \left| \int_0^b F_B(\beta) d\beta^n \right| \leq 2b^n.$$

Hence,

$$\begin{aligned}
 &\left| \sum_{n=0}^{\infty} \frac{(-T)^n E[B^n]}{n!} \right| \\
 &\leq \sum_{n=0}^{\infty} \frac{|(-T)^n| |E[B^n]|}{n!} \\
 &\leq 2 \sum_{n=0}^{\infty} \frac{T^n b^n}{n!} \\
 &= 2e^{Tb}
 \end{aligned}$$

which illustrates the convergence of  $\sum_{n=0}^{\infty} \frac{(-T)^n E[B^n]}{n!}$  for any  $T$ .

■ Theorem 1 and (5.1) explore only on the local effect: the epidemic spreads on average faster along a link in the heterogeneous case than the homogeneous case. However, if the infection probabilities of all the nodes are similar and the state of the each node (infected or not) is independent, each connected node pair would have a similar fraction of time when one node is infected whereas the other is susceptible, i.e. the period that allows epidemic to spread. In this case, the difference  $\chi(T)$ , where  $0 \leq \chi(T) < 1$ , in infection probability along a link within an arbitrary time  $T$  may indicate the difference in the fraction of infected nodes between the homogeneous and heterogeneous SIS in the metastable state. Both the heterogeneous infection rates and the heterogeneous network topology contribute to the heterogeneity in the infection probability of each node. When the recovery rate is low or equivalently the epidemic prevalence is high, however, the infection probabilities of the nodes tend to be similar. Hence,  $\chi(T)$  could suggest the difference in the fraction of infected nodes between the heterogeneous and homogeneous cases when the recovery rate is small. The larger the difference  $\chi(T)$  is, the smaller the average fraction  $y_{\infty}$  of infected nodes, in the metastable states of the heterogeneous SIS is. Equation (5.1), thus suggests that, the larger even-order moments of the infection rates lead to a smaller average fraction of infected nodes  $y_{\infty}$ , but the odd-order moments contribute in the opposite way. These theoretical results help us better understand our two observations in Fig. 5.1 and 5.2, when the recovery rates are small: (a) the average fraction  $y_{\infty}$  of infected nodes decreases with the increased variance, and (b) given the same variance, the average fraction  $y_{\infty}$  of infected nodes is lower if the third moment of the distribution is smaller.

### 5.3.3. THE LOG-NORMAL DISTRIBUTION VS. THE GAMMA DISTRIBUTION

To explore how fast  $y_{\infty}$  decays, we perform simulations with different recovery rates  $\delta$  and fit the curves of  $y_{\infty}$  vs. the variance  $\nu$ . We find that, as shown in Fig. 5.3, the relationship between the average fraction  $y_{\infty}(\nu)$  of infected nodes and the variance  $\nu$  can be fitted by a double-exponential function  $y_{\infty,L}(\nu) = c_1 e^{-c_2 \nu} + c_3 e^{-c_4 \nu}$  and a quadratic function  $y_{\infty,\Gamma}(\nu) = c_1 \nu^2 - c_2 \nu + c_3$ , when the infection rates follow log-normal and gamma distributions respectively. The coefficients  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$ , shown in Table 5.1, also suggest that, approximately,  $y_{\infty,L}$  decreases exponentially with the variance  $\nu$  much slower than the linear decrease of  $y_{\infty,\Gamma}$  when  $y_{\infty,\Gamma}$  is not close to 0.

Table 5.1: The coefficients of the fitting functions of  $y_{\infty}$  vs. the variance  $\nu$  for different infection-rate distributions under different recovery rates.

Dist.	$\delta$	$c_1$	$c_2$	$c_3$	$c_4$
$\text{Log} - \mathcal{N}$	3	0.098	0.045	0.099	0.28
	2	0.20	0.011	0.21	0.14
$\Gamma$	1	0.0011	0.055	0.67	N/A
	0.5	0.00085	0.053	0.83	

Besides the theoretical explanation as mentioned before, we explore further the physical interpretations of the difference in the fraction of infected nodes between the log-

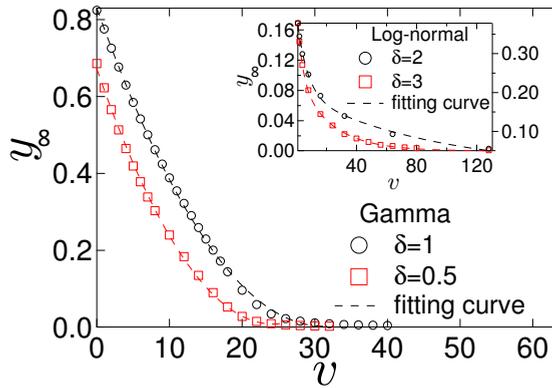


Figure 5.3: The average fraction  $y_\infty$  of infected nodes as a function of the variance  $\nu$  of infection rates following gamma distributions. The recovery rates  $\delta$  are different: 1 ( $\circ$ ) and 0.5 ( $\square$ ), and the dash lines are fitting curves. The simulations are on ER networks with average degree  $\langle k \rangle = 4$  and network size  $N = 10^4$ . The results are averaged over 1000 realizations. The inset contains the results about log-normal distributions.

normal and gamma distributed infection rates. We define  $r(\beta)$  as the ratio between the PDF of the log-normal and gamma distribution, *i.e.*  $r(\beta) = \frac{f_B(\beta; \mu, \sigma)}{f_B(\beta; k, \theta)}$ . Thus  $\lim_{\beta \rightarrow 0} r(\beta) = 0$  and  $\lim_{\beta \rightarrow \infty} r(\beta) = \infty$ . This reveals that if we set the same mean and variance (large) for both distributions, the log-normal distribution tends to generate a few extremely large values whereas the gamma distribution generates many extremely small values to produce the large variance.

Table 5.2: The percentiles of the log-normal and gamma distribution with the mean  $m_1 = 1$  and variance  $\nu = 16$

Percentiles	$\text{Log} - \mathcal{N}$	$\Gamma$
1 <sup>st</sup>	0.00483	$9.44 \times 10^{-32}$
2.5 <sup>th</sup>	0.00895	$2.20 \times 10^{-25}$
5 <sup>th</sup>	0.0152	$1.44 \times 10^{-20}$
10 <sup>th</sup>	0.0280	$9.44 \times 10^{-16}$
25 <sup>th</sup>	0.0779	$2.20 \times 10^{-9}$
50 <sup>th</sup>	0.243	$1.44 \times 10^{-4}$

In Table 5.2, we show the percentiles<sup>5</sup> of the two distributions with a large variance  $\nu = 16$ . In a group of random numbers generated by the gamma distribution, 25% of them are even smaller than  $2.2 \times 10^{-9}$ . The infection events driven by such small rates

<sup>5</sup>A percentile is a measure to indicate the value below which a given percentage of observations in a group of observations fall.

can hardly happen. However, in the infection rates generated by the log-normal distribution, even the first 1% smallest values are large enough to make possible infections. Hence, the gamma distribution effectively filters the network more than the log-normal distribution, and reduce the spread of the epidemic more. This interpretation is also consistent with the theoretical explanation of the influence of the third moment of a distribution. The same large variance can be introduced by the log-normal distribution via the possibility of generating a large value and by the gamma distribution via the high probability of generating extremely small values. However, the gamma distribution leads to a smaller third moments compared to the log-normal distribution and the small infection rates it generates effectively filter the network, reducing the epidemic spread.

## 5.4. LARGE RECOVERY RATES

We have shown that when the recovery rates are small, the i.i.d. heterogeneous infection rates retards the epidemic spreading and the larger variance of infection rates leads to a smaller average fraction of infected nodes. Moreover, we further explained the influence of the higher moments of the infection rate on epidemic spreading. In this section, we discuss how the heterogeneous infection rates influence the epidemic spreading when the recovery rate is large, thus, the epidemic is close to die out. As an example, we show the simulation results of the SF networks with the log-normal distributed infection rates. We find that, the heterogeneous infection rates may increase the probability that the epidemic spreads out when the recovery rate is large, though if the epidemic can spread out, the larger variance of infection rates still leads to a smaller average fraction of infected nodes in the metastable state.

We first employ the log-normal distribution for the heterogeneous infection rates and set the recovery rate  $\delta = 20$ . As shown in Fig. 5.4, though the average fraction  $y_\infty$  of infected nodes is close to 0 (due to the large recovery rate), we can observe that the larger variance may lead to a slightly larger average fraction  $y_\infty$  of infected nodes. However, the error bars (the standard deviation of the simulation results from different realizations) are large as compared to the average fraction of infected nodes. This is due to the fact that when the epidemic is close to die out on average, i.e. when  $\delta = 20$ , the epidemic dies out in some iterations of the simulations but spreads out with a nonzero fraction of infected nodes in the metastable state in the others.

Fig. 5.5 shows the percentage  $p^*$  ( $\in [0, 1]$ ) of the spread-out realizations in all realizations and the average fraction  $y_\infty^*$  of infected nodes in these nonzero-infection realizations as a function of the variance of the infection rates. Here the the simulations are on SF networks with the size  $N = 10^4$  and the exponent  $\lambda = 2.5$ . Clearly, the average fraction of infected nodes obtained by averaging that in all realizations is  $y_\infty = p^* y_\infty^*$ . We find that, in all nonzero-infection realizations, the average fraction  $y_\infty^*$  of infected nodes still decreases as the variance of the infection rates increases. The average fraction  $y_\infty$  of infected nodes obtained from all realizations may increase as the variance of the infection rates increases, because the percentage  $p^*$  of nonzero-infection realizations increases when the variance of the infection rates is small and increases. Hence, the heterogeneous infection rates may enhance the probability that the epidemic spreads out. This can be explained as follows: the heterogeneous infection rates and the hubs in scale-free networks enable those links with a large infection rate to form a connected subgraph,

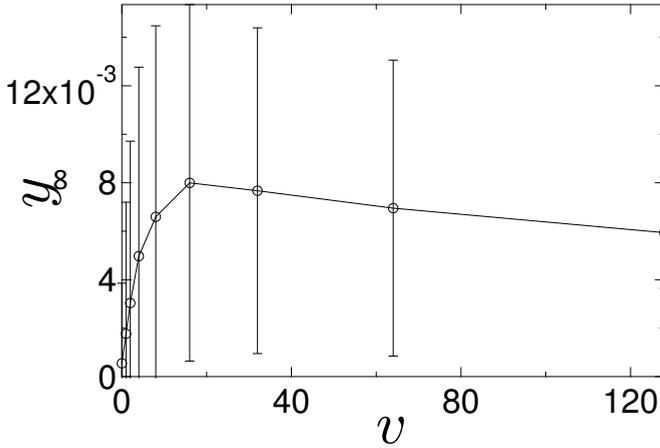


Figure 5.4: The average fraction  $y_\infty$  of infected nodes as a function of the variance  $\nu$  of infection rates following the log-normal distribution, and the recovery rate  $\delta = 20$ . The simulations are on SF networks with the exponent  $\lambda = 2.5$  and the network size  $N = 10^4$ . The results are averaged over 1000 realizations, and the error bars are the standard deviations of the results in different realizations.

allowing the epidemic to spread out. However, when the variance  $\nu$  is large and further increases, as shown in Fig. 5.5, the fraction of non-zero infection realizations decreases. This is because, a large variance  $\nu$  of the infection rates produces fewer large infection rates, prohibiting the formation of a connected subgraph with high infection rates that allows the epidemic to spread. However, the average fraction of infected nodes of the nonzero-infection realizations tend to decrease with the variance or heterogeneity of the infection rates.

If we increase the recovery rate to ensure that the epidemic dies out in the homogeneous case, *i.e.* the effective infection rate is below the epidemic threshold  $\tau_c$  in the classic SIS model, we obtain the same conclusions: the average fraction  $y_\infty^*$  of infected nodes in nonzero-infection realizations (if exist) always decreases as the variance of the infection rates increases, and the heterogeneous infection rates may increase the probability that the epidemic spreads out.

We further compare the simulation results between the log-normal and gamma distributions. As shown in Fig. 5.6(a), the average fraction  $y_\infty^*$  of infected nodes in nonzero-infection realizations is larger when the infection rates follow the log-normal distribution than the gamma distribution. This observation is consistent with our previous observations and conclusions as illustrated in Section 5.3, when the variances of the infection rates are the same, the larger third moments of the infection rates lead to the more severe infection. However, as shown in Fig. 5.6(b), when the variance of the infection rates is small, the percentage  $p_T^*$  of the nonzero-infection realizations is larger in the case of the gamma distributed infection rates than the percentage  $p_L^*$  of the nonzero-infection realizations in the case of the log-normal distributed infection rates. More-

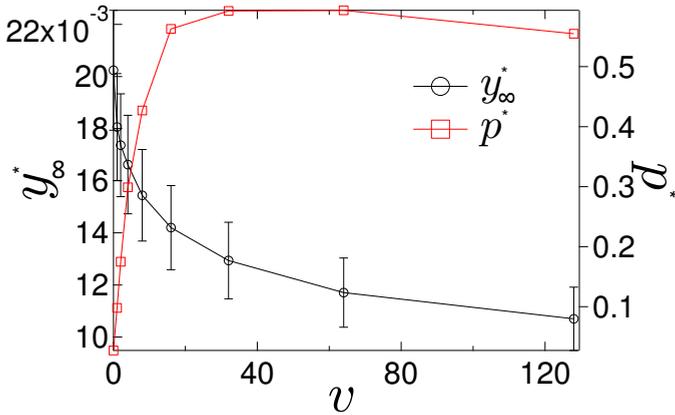


Figure 5.5: The average fraction  $y_{\infty}^*$  ( $\circ$ ) of infected nodes in the nonzero-infection realizations and the percentage  $p^*$  ( $\square$ ) of the nonzero-infection realizations as a function of the variance  $v$  of infection rates. The infection rates follow the log-normal distribution and the recovery rate  $\delta = 20$ . The simulations are on SF networks with the exponent  $\lambda = 2.5$  and the network size  $N = 10^4$ . The results are averaged over 1000 realizations, and the error bars are the standard deviations of the results in different realizations.

over, as the variance of the infection rates is relatively large (for example, around 30 in Fig. 5.6(b)) and increases,  $p_{\Gamma}^*$  decreases faster than  $p_L^*$ , and  $p_{\Gamma}^*$  could be smaller than  $p_L^*$  if the variance is large enough. Given a network and a large recovery rate, more large infection rates lead to a higher probability that the epidemic can spread out. As in Section 5.3.3, we can explain the observations in Fig. 5.6(b) by exploring the percentiles of the log-normal and gamma distributions with the mean 1 in Table 5.3, where two values (16 and 128) of the variance are employed as examples. When the variance is 16, there are more large values in a group of random numbers generated by the gamma distribution than the log-normal distribution; however, when the variance increases to 128, though the first 1% largest values of the gamma distribution are still larger than those of the log-normal distribution, there are more large values in the group of the log-normal random numbers. Hence, with the same small variance, the gamma distributed infection rates contribute more to the survival of the epidemic than the log-normal distributed infection rates, whereas with the same large variance, the log-normal distributed infection rates may lead to a higher probability that the epidemic spreads out.

We observe the same in ER networks though not shown here. Moreover, the links with i.i.d. large infection rates are more likely to form a subgraph in SF networks than in ER networks, because of the existence of the nodes with large degrees in SF networks. Hence, with the similar value of the average fraction  $y_{\infty}^*$  of infected nodes in the nonzero-infection realizations, we find that the percentage  $p^*$  of the nonzero-infection realizations is much smaller in ER networks than SF networks.

We further consider an extreme case of SF networks – the star network: one central node  $n_0$  connects with all the other  $m$  ( $m \gg 1$ ) side nodes  $n_i$  ( $i = 1, 2, \dots, m$ ), and there is no link between any pair of the side nodes. By designing a specific distribution of the

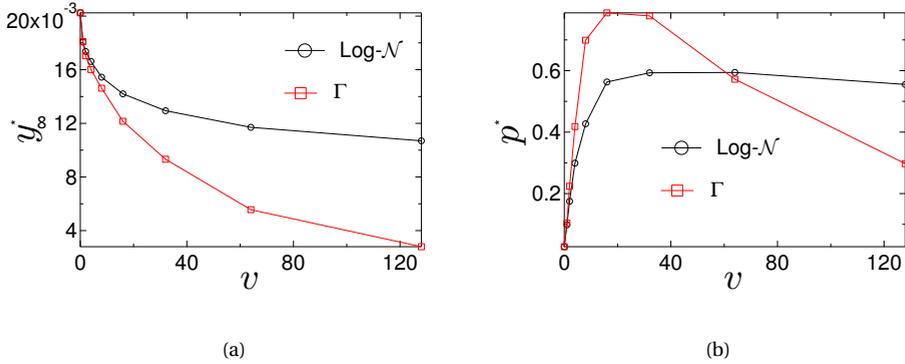


Figure 5.6: (a) The average fraction  $y_\infty^*$  of infected nodes in the nonzero-infection realizations and (b) The percentage  $p^*$  of the nonzero-infection realizations in all realizations as a function of the variance  $\nu$  of the infection rates which follow the gamma ( $\circ$ ) and log-normal ( $\square$ ) distribution.

Table 5.3: The percentiles of the log-normal and gamma distributions with the mean  $m_1 = 1$  and variance  $\nu = 16$  and 128

	$\text{Log-}\mathcal{N}$ $\nu = 16$	$\Gamma$ $\nu = 16$	$\text{Log-}\mathcal{N}$ $\nu = 128$	$\Gamma$ $\nu = 128$
99 <sup>th</sup>	12.1936	19.9409	15.1178	24.2306
98 <sup>th</sup>	7.7225	12.9981	8.2133	5.7424
97 <sup>th</sup>	5.7697	9.3923	5.6176	1.5509
96 <sup>th</sup>	4.6209	7.1783	4.2063	0.4140
95 <sup>th</sup>	3.8751	5.6325	3.3260	0.1507

heterogeneous infection rates, we can always give a value of the recovery rate  $\delta$  so that the epidemic spreads out with the heterogeneous infection rates but dies out with the corresponding homogeneous infection rates in a finite-size star network. In the classic model, the epidemic threshold of a star network is  $\tau_c = \frac{\beta}{\delta} = \frac{1}{\sqrt{m}}$  [121]. If we set the homogeneous infection rate  $\beta = 1$  and the recovery rate  $\delta = \sqrt{m} + \epsilon$ , where  $\epsilon$  is a positive but small constant number, then the epidemic dies out. With the same recovery rate, we set the heterogeneous infection rate with the distribution  $Pr[B = 2 - \epsilon_1] = Pr[B = \epsilon_1] = 0.5$ , where  $\epsilon_1$  is again a small and positive constant number, thus the average infection rate  $E[B] = 1$ . We now look at the subgraph which is composed of the central node and approximately  $\frac{m}{2}$  side nodes connected to the central node with infection rate  $\beta_{sub} = 2 - \epsilon_1$ . The effective infection rate is  $\tau_{sub} = \frac{\beta_{sub}}{\delta} = \frac{2 - \epsilon_1}{\sqrt{m} + \epsilon_1} \approx \frac{2}{\sqrt{m}} > \frac{1}{\sqrt{m/2}} \approx \tau_{c,sub}$ , where  $\tau_{c,sub}$  is the epidemic threshold of the subgraph. Hence, with the same recovery rate and the same average infection rates, the epidemic dies out in the homogeneous case but spreads out in the aforementioned heterogeneous case.

## 5.5. REAL-WORLD NETWORKS

As mentioned in Section 5.2.2, the interaction frequency between two nodes in a real-world network has been considered as the infection rate between the pair of nodes. In this section, we choose two real-world networks as examples to illustrate how their heterogeneous infection rates affect the spread of SIS epidemics on these networks. The heterogeneous infection rates from the datasets are normalized by the average so that the average is 1. We compare the average fraction of infected nodes in the metastable state of the two networks in the 3 scenarios: 1) each network is equipped with its normalized original heterogeneous infection rates (hetero- $\beta$ ) as given in the dataset; 2) each network is equipped with the infection rates in the normalized original dataset but randomly shuffled (shuffled- $\beta$ ); 3) each network is equipped with a constant infection rate (homo- $\beta$ ) which equals to the average infection rate of the normalized original infection rates as given in the datasets. The heterogeneous infection rates in each network described in Scenario 1 are possibly correlated. For example, the infection rate of a link may depend on the degrees of the two ending nodes of this link. The shuffling in Scenario 2 effectively removes the correlation if it exists, and the infection rates in Scenario 3 are homogeneous as in the classic SIS model. Our objective is to explore the relation between the infection rates and average fraction of infection in these 3 scenarios for both networks to verify our previous findings.

The first network is the airline network where the nodes are the airports, the link between two nodes indicates that there's at least one flight between these two airports, and the infection rate along a link is the number of flights between the two airports. We construct this network and its infection rates from the dataset of openFlights<sup>6</sup>. The other one is the co-author network, where the nodes are the authors of papers, the link represents that the two corresponding authors have at least one collaborated paper, and the infection rate is the collaboration frequency[87].

Besides the infection rates, the network topology may as well influence the spread of

<sup>6</sup><http://openflights.org/data.html>

SIS epidemics. We explore the most fundamental network feature of the two networks: the degree distributions which are shown in Fig. 5.7. We can see that the degree distributions of the airline network and co-author network approximately follow a power law with the slope  $\lambda = 1.5$  and  $2.5$  respectively. Hence, the degree distributions of the two networks influence the spread of epidemics in a similar way. More details of the two networks are listed in Table 5.4. Note that we normalized the infection rates of each network by its mean so that the average rate is 1.

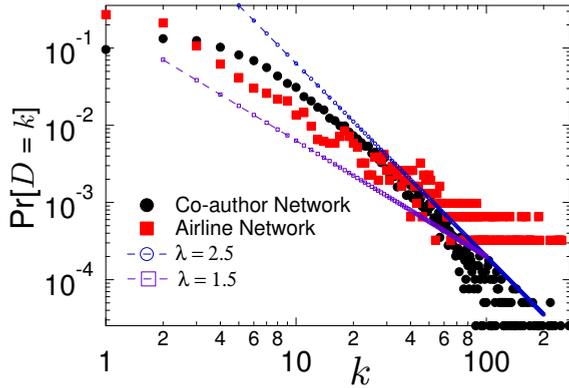


Figure 5.7: The degree distributions of the airline (■) and co-author network(●) can be approximately fitted by the power law distribution with a slope  $\lambda = 1.5$  and  $2.5$  respectively.

Table 5.4: The number of nodes, number of links, variance of infection rates and range of infection rates in the two networks.

Name	Nodes	Links	Variance	Range
Airline	3071	15358	0.5560	[0.2383, 11.0626]
Co-author	39577	175692	3.0566	[0.0678, 90.4625]

The distributions of the infection rates from the two networks are shown in Fig. 5.8(a) and 5.8(b). We find that, approximately the infection rates of the airline network are exponentially distributed, whereas those of the co-author network follow a log-normal distribution. Both of the two datasets support our previous choices of the infection-rate distribution.

### 5.5.1. SMALL RECOVERY RATES

We first consider the small recovery rates, with which the epidemic does not die out in any realizations. In this chapter, we assume that the infection rates are i.i.d. which corresponds to Scenario 2. As shown in Fig. 5.9(a) and 5.9(b), the average fraction  $\gamma_\infty$  of infected nodes in Scenario homo- $\beta$  is always larger than that in Scenario shuffled- $\beta$ , which

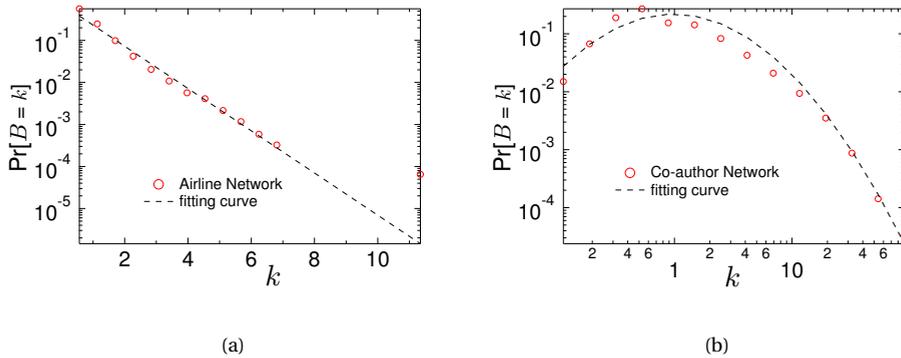


Figure 5.8: The distribution of the infection rates from real-world networks: (a) airline network and (b) co-author network. In each figure, the distribution ( $\circ$ ) and fitting curve (dash line) are shown. The fitting curves are exponential and log-normal distributions in (a) and (b) respectively.

confirms our conclusion that the heterogeneity of infection rates on average retards the contagion processes of epidemics, when recovery rates are not very large. Moreover, we find that the reduction  $y_{\infty, \text{homo-}\beta} - y_{\infty, \text{shuffled-}\beta}$  is larger in the co-author network, which has a larger variance of infection rates, than that in the airline network<sup>7</sup>. This observation verifies our conclusion that, the larger the variance of the infection rates is, the smaller  $y_{\infty}$  is. Compared to the independent infection rates in the case shuffled- $\beta$ , the possibly correlated infection rates in the case hetero- $\beta$  can further decrease (in e.g. the airline network) or increase (in e.g. the co-author network) the average fraction of infected nodes. This observation points out a new challenging question: what is the influence of such correlated heterogeneous infection rates on the SIS epidemics.

### 5.5.2. LARGE RECOVERY RATES

As shown in Fig. 5.10, when the recovery rate increases and the effective infection rate is close to the epidemic threshold, the average fraction  $y_{\infty}$  of infected nodes in the Scenario hetero- $\beta$  becomes mostly larger than that in the other two scenarios. Besides that, it is still consistent with our previous conclusion that if  $y_{\infty, \text{homo-}\beta} \neq 0$ , then  $y_{\infty, \text{homo-}\beta} > y_{\infty, \text{shuffled-}\beta}$ . Moreover, in the co-author network, we observe that when the recovery rate  $\delta = 40$ ,  $y_{\infty, \text{shuffled-}\beta} > y_{\infty, \text{homo-}\beta} = 0$ . However, in the airline network, we cannot observe that  $y_{\infty, \text{shuffled-}\beta} > y_{\infty, \text{homo-}\beta}$  with any selected recovery rate, and this may be because of the small variance of the infection rates. Hence, the observations verify our conclusions that if the epidemic spreads out with the homogeneous infection rates, then the overall infection is always more severe than that with the heterogeneous infection rates (i.i.d. and with the same mean as the homogeneous infection rate); however, the heterogeneous infection rate may contribute to the survival of the epidemic.

<sup>7</sup>We assume that the two networks have a similar topology, since they have a similar degree distribution as shown in Fig. 5.7

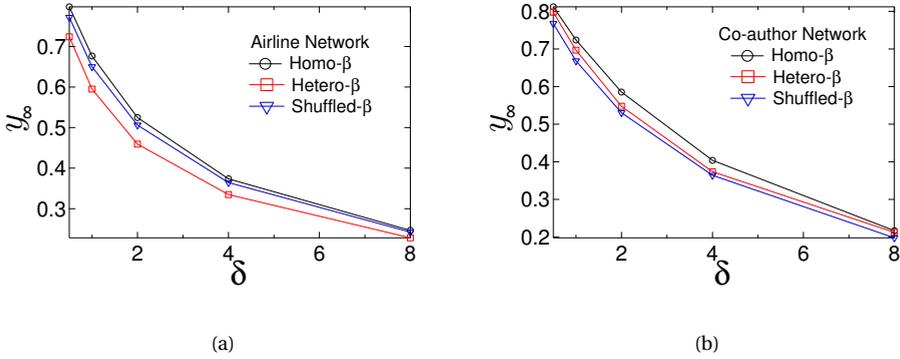


Figure 5.9: The average fraction  $y_\infty$  of infected nodes as a function of the recovery rate  $\delta$ . The networks are from real-world: (a) airline network and (b) co-author network. In each figure, the SIS model with homogeneous (○), original heterogeneous (□) and shuffled heterogeneous (▽) infection rates are compared.

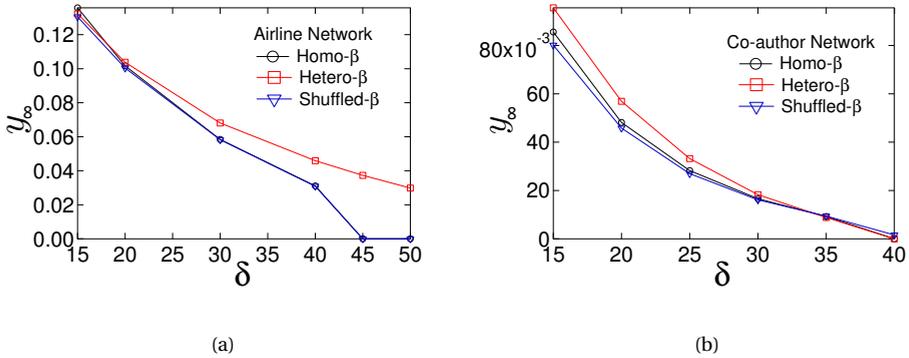


Figure 5.10: The average fraction  $y_\infty$  of infected nodes as a function of the recovery rate  $\delta$ . The recovery rate is very large so that the effective infection rate is close to the epidemic threshold. The networks are from real-world: (a) airline network and (b) co-author network. In each figure, the SIS model with homogeneous (○), original heterogeneous (□) and shuffled heterogeneous (▽) infection rates are compared.

## 5.6. DISCUSSIONS

In summary, we illustrate with simulations, theoretical analysis and physical interpretations that, when the recovery rate is small, the heterogeneity of infection rates on average retards the virus spread and whereas the larger even-order moments of the infection rates tend to lead to a smaller  $y_\infty$ , the odd-order moments contribute in the other way around; when the recovery rate is large so that the epidemic may die out, the heterogeneous infection rates may enhance the probability that the epidemic spread out. We also verify the influence of the heterogeneity of infection rates on virus spread in real-world networks. Our work reveals that the higher moments, especially the variance, of the infection rates may evidently affect the epidemic spread, even far more seriously than intuitively expected. Our finding implies that real-world heterogeneous epidemic spread may not be as severe as the classic homogeneous SIS model predicts, but the heterogeneous epidemic may not be as easy as the homogeneous SIS model indicates to die out.

In this chapter, we have focused on the Markovian SIS where the time for an infected node  $i$  to infect a susceptible neighbor  $j$  is an exponential random variable with rate  $\beta_{ij}$ . Theorem 1 can be extended to Non-Markovian SIS models with heterogeneous infection rates where the infection time between a neighboring infected susceptible node pair  $(i, j)$  with average  $1/\beta_{ij}$  follows a distribution other than the exponential distribution. Such extension to Non-Markovian SIS models is possible if  $1 - F(\tau; \beta)$  the probability that the infection time is larger than  $\tau$  when the average infection time is  $1/\beta$  is a convex function of  $\beta$ .

The time for an infected node to infect a susceptible neighbor is more in depth and detailed information. Infection time measurement becomes possible though in general is still challenging. For example, in the experiments of the epidemic in the plant population, the infection time can be measured. As more such datasets become available, it would be interesting to tackle a new direction: what is the influence of the heterogeneous infection time on viral spreading?



# 6

## EPIDEMIC SPREADING WITH DEGREE-CORRELATED HETEROGENEOUS INFECTION RATES

*The infection rate between a pair of nodes, which may depend on e.g. their interaction frequency, is usually heterogeneous and even correlated with their nodal degrees in the contact network. In this chapter, we aim to understand how such correlated heterogeneous infection rates influence the epidemic spreading on different network topologies. Motivated by real-world datasets, we propose a correlated heterogeneous Susceptible-Infected-Susceptible (CSIS) model which assumes that the infection rate  $\beta_{ij} (= \beta_{ji})$  between node  $i$  and  $j$  is correlated with the degree of the two end nodes:  $\beta_{ij} = c(d_i d_j)^\alpha$ , where  $\alpha$  indicates the strength of the correlation between the infection rates and nodal degrees, and  $c$  is selected so that the average infection rate is 1 in this work. In order to understand the effect of such correlation on epidemic spreading, we consider as well the corresponding uncorrected but still heterogeneous infection rate scenario as a reference, where the original correlated infection rates in our CSIS model are shuffled and reallocated to the links of the same network topology. We compare these two scenarios in the average fraction of infected nodes in the metastable state on ER and SF networks with a similar average degree. We find that, when the recovery rate is small, the negative correlation is more likely to help the epidemic spread and the positive correlation prohibit the spreading; as the recovery rate increases to be larger than a critical value, the positive but not negative correlation tends to help the spreading. Our findings are further analytically proved in a wheel network (one central node connects with each of the nodes in a ring) and validated on real-world networks with correlated heterogeneous interaction frequencies.*

---

This chapter have been published as: B. Qu and H. Wang, SIS Epidemic Spreading with Correlated Heterogeneous Infection Rates, *Physica A Statistical Mechanics & Its Applications*, 2017, 472(23–24):4543-4548.

## 6.1. INTRODUCTION

**B**IOLOGICAL, social and communication systems can be represented as networks by considering the system components or individuals as nodes and the interactions or relations in between nodes as links. Viral spreading models have been used to model processes e.g. the propagation of information and epidemics on such networks or complex systems [33, 91, 92, 95]. The Susceptible-Infected-Susceptible (SIS) model is one of the most studied models. In the SIS model, at any time  $t$ , the state of a node is a Bernoulli random variable, where  $X_i(t) = 0$  represents that node  $i$  is susceptible and  $X_i(t) = 1$  if it is infected. Each infected node infects each of its susceptible neighbors with an infection rate  $\beta$ . The infected node can be recovered to be susceptible again with a recovery rate  $\delta$ . Both infection and recovery processes are independent Poisson processes. The ratio  $\tau \triangleq \beta/\delta$  is called effective infection rate. When  $\tau$  is larger than the epidemic threshold  $\tau_c$ , the epidemic spreads out with a nonzero fraction of infected nodes in the metastable state. The average fraction of infected nodes  $y_\infty$  in the metastable state, ranging in  $[0, 1]$ , indicates how severe the influence of the virus is: the larger the fraction  $y_\infty$  is, the more severely the network is infected.

In the classic SIS model, both the infection and recovery rates are assumed homogeneous, i.e. the infection rates are the same for all infected-susceptible node pairs and the recovery rates are the same for all infected nodes. However, the infection rates which can be reflected from, for example, the interaction frequencies, between nodes in real-world networks are usually heterogeneous and even dependent on the properties of the nodes. For examples, the number of flights between different pairs of airports in a month are different in the airline transportation network, and the number of collaborated papers between different pairs of authors in a year vary in a co-author network [11, 77]. The interaction frequency is found to be correlated with the nodal degrees in e.g. airline transportation network and metabolic network [11, 77, 81]. Hence, we aim to understand the effect of correlated infection rate on the viral spreading in this work. We propose a correlated heterogeneous SIS (CSIS) model, in which the recovery rates are homogeneous but the infection rate  $\beta_{ij}(= \beta_{ji})$  between node  $i$  and  $j$  is correlated with their degrees  $d_i$  and  $d_j$  in the way:

$$\beta_{ij} = c(d_i d_j)^\alpha \quad (6.1)$$

where  $\alpha$  indicates the strength of the correlation and  $c$  is a constant to control the average infection rate to 1. The correlation strength  $\alpha \approx 0.5$  in the network of airports (both in US [11, 81] and China [77]) and  $\alpha \approx 0.8$  in the metabolic network [81].

This paper aims to understand the effect of correlation between the infection rates and nodal degrees on viral spreading. Motivated by real-world networks, we consider the generic case where the heterogeneous infection rates are correlated with nodal degrees as described by (6.1) and the network topology is as well heterogeneous. We consider also the corresponding uncorrelated heterogeneous infection rates scenario, where the correlated infection rates are shuffled and randomly assigned to all the links as a reference scenario. For the heterogeneity of the network topology, we consider the different broadness of the degree distribution, i.e. Erdős-Rényi (ER) and scale-free (SF) networks. We explore how the positive or negative correlation between the infection rates and nodal degrees influences the epidemic spread on the networks with different heterogeneities of the topology.

A few recent papers have taken into account the heterogeneous infection rates in the SIS model. We studied the SIS model with identically and independently distributed (i.i.d.) infection rates and found that independent of the underlying network topology, though the heterogeneous infection rates may increase the probability that the epidemic spreads out when the recovery rate is large, the overall infection decreases as the variance of infection rates increases if the epidemic spreads out [100]. Buono et al. [23] considered a specific distribution of infection rates and observed slow epidemic extinction phenomenon. Preciado et al. [97] discussed how to choose the infection and recovery rates from given discrete sets to let the virus die out. Yang and Zhou [135] gave an edge-based mean-field solution of the epidemic threshold in regular networks (the degrees of all nodes are the same) with i.i.d. heterogeneous infection rates following a uniform or power-law distribution. Fu et al. [42] analytically studied the epidemic threshold when the infection rates are as a piecewise linear function of degrees by their mean-field method. Van Mieghem and Omic [124] developed the N-intertwined mean field approximation (NIMFA) [121] for heterogeneous SIS model, and discussed the bounds of epidemic threshold and the convexity of the steady-state infection probability of each node as function of the recovery rate. Heterogeneous infection rates have also been considered in interconnected networks, allowing component networks and their interconnections to have a different but constant infection rate [105, 128]. In this paper, we employ the continuous-time simulation of the exact SIS model instead of the mean-field methods and introduce a general correlation between the nodal degree and the infection rate which is motivated by real-world datasets to study the influence of the correlation on the epidemic spreading.

## 6.2. PRELIMINARY

In this section, we first introduce the construction of network models and the heterogeneous infection rates which are considered in our CSIS model. We then introduce the continuous-time simulation which is the main approach in this work. Finally, we introduce the mean-field approximation of the SIS and CSIS model which will be further used in our theoretical analysis of the CSIS model on the wheel network in Section 6.4.

### 6.2.1. THE INFECTION RATES

Given the network topology, we build two heterogeneous infection-rate scenarios: 1) the correlated infection rates and 2) the uncorrelated or shuffled infection rates. In the scenario of correlated infection rates, we assume that  $\beta_{ij} = c(d_i d_j)^\alpha$  where  $\alpha$  indicates the correlation strength. We selected the constant  $c$  such that the average infection rate is 1, whereas we consider different values of homogeneous recovery rates. In this case, the infection rate of each link is determined by the given network topology and  $\alpha$ . In the scenario of uncorrelated infection rates, we shuffle the infection rates from all the links as generated in the first scenario and redistribute them randomly to all the links. In this way, we keep the distribution of infection rates but effectively remove the correlation between the infection rates and nodal degrees. The homogeneous infection rate is a special case of our heterogeneous infection rate construction where  $\alpha = 0$ . Clearly, in a homogeneous network where all the nodes have the same number of neighbors, the

infection rates are homogeneous in both scenarios for any  $\alpha$ .

As examples, we show the distribution of the heterogeneous infection rates when  $\alpha = 0.5$  (the positive correlation) and  $\alpha = -0.5$  (the negative correlation) for both ER and SF networks in Fig. 6.1. We find that the positive correlation between the infection rates and the nodal degrees leads to a heavy-tailed distribution of the infection rate in SF networks. However, the negative correlation in SF networks and both kinds of correlation in ER networks only generate the infection rates distributed in a small range.

A positive  $\alpha > 0$  (or negative  $\alpha < 0$ ), suggests a positive (or negative) correlation between the infection rates and nodal degrees. Too large or too small values of  $\alpha$  could not be realistic. For example, [11, 77, 81] suggest that  $\alpha$  is around 0.5 or 0.8 in their datasets, and we also find  $\alpha \approx 0.14$  and  $-0.12$  respectively in two real-world dataset as described in Section 5. Hence, we focus mainly on the range of  $\alpha \in [-1, 1]$  in this paper, and discuss the extreme case when the absolute value of  $\alpha$  is large in Section 6.3.2.

### 6.2.2. THE SIMULATIONS

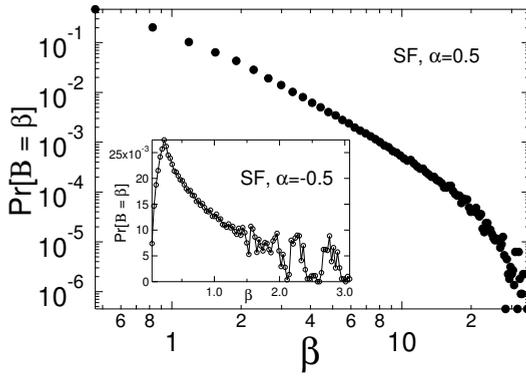
In this chapter, we perform the continuous-time simulations of the CSIS model on both ER networks and SF networks (the heterogeneity of the network topology is thus taken into account) with  $N = 10000$  nodes. Given a network topology, a recovery rate  $\delta$  and a value of  $\alpha$ , we carry out 100 iterations. In each iteration, we generate the heterogeneous infection rates as described in (6.1) for the scenario of the correlated infection rates and shuffle them for the scenario of uncorrelated infection rates. Initially, 10% of the nodes are randomly infected. Then the infection and recovery processes of SIS model are simulated until the system reaches the metastable state where the fraction of infected nodes is unchanged for a long time. The average fraction  $y_\infty$  of infected nodes is obtained over 100 iterations for both scenarios of the correlated and uncorrelated infection rates. Moreover, for simplicity, we use  $y_{\infty,c}$  and  $y_{\infty,u}$  to denote the average fraction of infected nodes in the scenarios of correlated and uncorrelated infection rates respectively.

## 6.3. EFFECT ON THE AVERAGE FRACTION $y_\infty$ OF INFECTED NODES

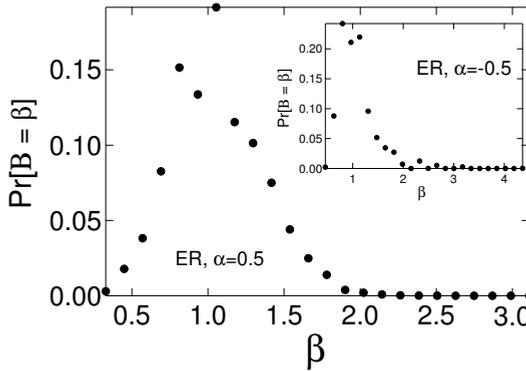
As mentioned, the average fraction  $y_\infty$  of infected nodes in the metastable state indicates how severe the network is infected. In this section, we explore how the average fraction  $y_\infty$  of the infected nodes depends on the parameter  $\alpha$  in both of the two scenarios: correlated and uncorrelated infection rates. We mainly consider the influence of the correlation between the infection rates and nodal degrees on epidemic spreading when the recovery rate varies and the absolute value of  $\alpha$  is in the range  $[-1, 1]$ . The difference of the influence between ER and SF networks is also discussed. Then we briefly describe the influence of the correlation in an extreme case when the absolute value of  $\alpha$  is much larger. In this case, the influence of the correlation is independent from the value of the recovery rate.

### 6.3.1. REALISTIC CASES: $\alpha \in [-1, 1]$

In this subsection, we first intuitively explain that how the correlation influences the epidemic spreading as the recovery rate  $\delta$  and the parameter  $\alpha$  vary. To support our explanations, we then define an intermediate quantity and illustrate the effect of the



(a)



(b)

Figure 6.1: The distribution of the heterogeneous infection rates for (a) SF networks and (b) ER networks, where the parameter  $\alpha = 0.5$  in the main figures and  $\alpha = -0.5$  in the insets.

correlation when the recovery rate is small and large respectively.

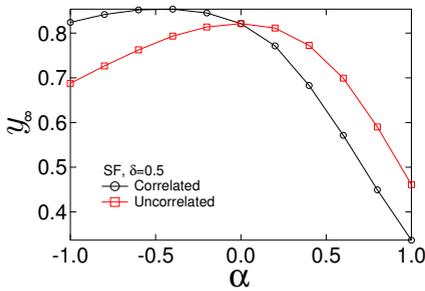
The average fraction  $y_\infty$  of infected nodes as a function of the scale parameter  $\alpha$  in both ER and SF networks are shown in Fig. 6.2. We employ different values of the recovery rate  $\delta$  to illustrate the influence of the correlation on the epidemic spreading over different range of recovery rates, i.e. different prevalence of the epidemic.

Our previous work [100] explored the SIS model with i.i.d. heterogeneous infection rates and showed that the average fraction  $y_\infty$  of infected nodes tends to decrease as the variance of the i.i.d. infection rates increases. In this chapter, the uncorrelated infection rates can be considered as being independent. In this case, our previous results can be applied to derive the relationship between the average fraction  $y_\infty$  of infected nodes and the parameter  $\alpha$  as shown in Fig. 6.2: as  $\alpha$  increases when  $\alpha > 0$  or as  $\alpha$  decreases when  $\alpha < 0$ , the variance of the infection rates increases, hence  $y_\infty$  decreases in the scenario of uncorrelated infection rates. In the scenario of correlated infection rates, though the peaks of  $y_\infty$  are not at  $\alpha = 0$  for both types of networks,  $y_\infty$  roughly decreases as the absolute value of  $\alpha$  increases. This is because, as the absolute value of  $\alpha$  increases, large infection rates are assigned to a small number of links, limiting the spread of the epidemic.

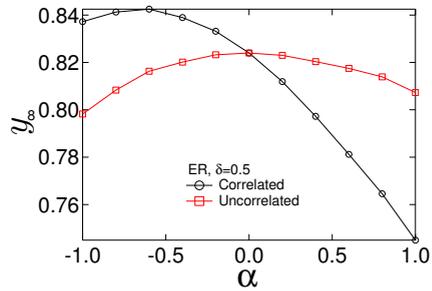
From Fig. 6.2, we find that 1) the negative correlation ( $\alpha < 0$ ) between the infection rates and the degrees tends to enhance the epidemic spreading compared to the uncorrelated infection-rate scenario ( $y_{\infty,c} > y_{\infty,u}$ ) when the recovery is small, but prohibit the spreading ( $y_{\infty,c} < y_{\infty,u}$ ) when the recovery rate is large; 2) the positive correlation ( $\alpha > 0$ ) tends to enhance the epidemic spreading ( $y_{\infty,c} > y_{\infty,u}$ ) when the recovery is large, but prohibit the spreading ( $y_{\infty,c} < y_{\infty,u}$ ) when the recovery rate is small.

In the scenario of uncorrelated infection rates, the parameter  $\alpha$  determines only the distribution of the i.i.d. infection rates in a network, and the infection rate between any pair of nodes is independent from their degrees. Compared to the scenario of uncorrelated infection rates, a positive correlation between the infection rates and nodal degrees ensures that the infection rates between nodes with larger degrees are also larger, whereas a negative correlation suggests the other way around. Intuitively, we may think that if the nodes with larger degrees can be infected by larger infection rates, the infection probability of those nodes are higher and those nodes can more effectively infect their neighbors as well. Hence, the positive correlation between the infection rates and nodal degrees seems to contribute to the epidemic spreading. This is indeed the case when the recovery rate  $\delta$  is large, i.e. the prevalence of the epidemic is low.

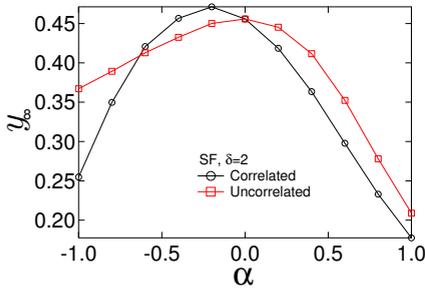
In contrast to this intuition, we observe that the positive correlation actually tends to prohibit the epidemic spreading when the recovery rate  $\delta$  is small. This can be explained as follows: when the recovery rate  $\delta$  is small, i.e. the prevalence is high, the infection probabilities of the large-degree nodes are already high, then the increment of the infection rates between the large-degree nodes may not significantly increase the infection probabilities of these nodes, and thus the infection probabilities of their neighbors may not be significantly increased. However, the negative correlation between the infection rates and nodal degrees leads to the higher infection rates between the small-degree nodes and effectively enhances the probabilities of the small-degree nodes compared to the scenario of the uncorrelated infection rates. Though the infection probabilities of large-degree nodes decrease in this case, the large amount of small-degree nodes en-



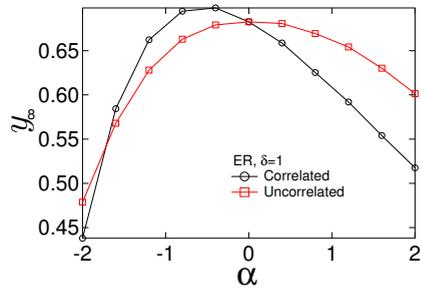
(a)



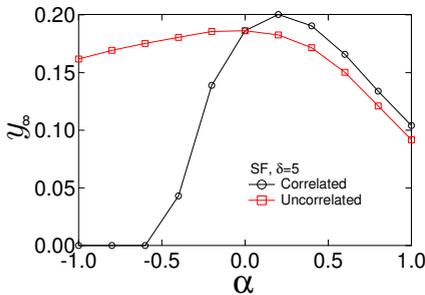
(b)



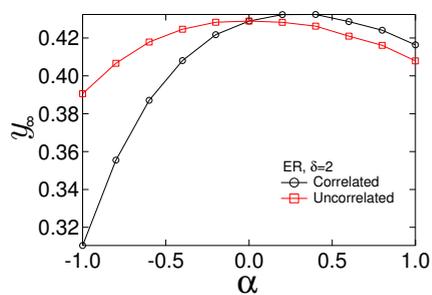
(c)



(d)



(e)



(f)

Figure 6.2: The average fraction  $y_\infty$  as a function of  $\alpha$  for (a) SF networks with the recovery rate  $\delta = 0.5$ , (b) ER networks with the recovery rate  $\delta = 0.5$ , (c) SF networks with the recovery rate  $\delta = 2$ , (d) ER networks with the recovery rate  $\delta = 1$ , (e) SF networks with the recovery rate  $\delta = 5$  and (f) ER networks with the recovery rate  $\delta = 2$  in both scenarios of correlated ( $\circ$ ) and uncorrelated ( $\square$ ) infection rates.

sures that the overall infection is on average enhanced.

To support our explanations, we define  $y_\infty^{(d)}$  as the average infection probability of the nodes with degree  $d$ , and  $y_{\infty,c}^{(d)}$  and  $y_{\infty,u}^{(d)}$  as that for the scenario of correlated infection rates and uncorrelated infection rates respectively. We show  $y_\infty^{(d)}$  as a function of the degree  $d$  when the recovery rate is small,  $\delta = 0.5$ , and the correlation parameter  $\alpha = -0.6$  in Fig. 6.3 (the main figures). We find that such a negative correlation ( $\alpha = -0.6$ ) between the infection rates and nodal degrees indeed decreases the infection probabilities of large-degree nodes, but the infection probabilities of the small-degree nodes are also significantly lifted. Furthermore, the number of small-degree nodes is much larger than that of large-degree nodes in SF networks, and those are similar in ER networks. To illustrate the combined effect above of the two aspects, we define  $\eta(d)$  as (6.2), the product of the probability that a node has the degree  $d$  and the difference between the average infection probability of the nodes with the degree  $d$  in the scenarios of correlated and uncorrelated infection rates:

$$\eta(d) = (y_{\infty,c}^{(d)} - y_{\infty,u}^{(d)})\Pr[D = d] \quad (6.2)$$

and  $y_{\infty,c} - y_{\infty,u} = \sum_{d=1}^{d=N-1} \eta(d)$ . Note that a positive  $\eta(d)$  always indicates that  $y_{\infty,c} > y_{\infty,u}$ , i.e. the correlation lifts the infection probability of nodes with the degree  $d$  compared to the scenario of the uncorrelated infection rates. As in the insets of Fig. 6.3, we plot  $\eta(d)$  as a function of the degree  $d$  for both ER and SF networks. More plots<sup>1</sup> of  $\eta$  as a function of the degree  $d$  are shown in Fig. 6.4 where the cases with small degrees are shown in the main figures and those with relatively large degrees are shown in the insets. We find that in both networks the value of  $\eta$  is significantly large for the small-degree nodes and contributes more to a higher prevalence of the epidemic when the recovery rate is small and the correlation is negative as shown in Fig. 6.3, Fig. 6.4(c) and Fig. 6.4(d). The observation is consistent with our explanation about why the negative correlation tends to help the epidemic spreading when the recovery rate is small. In contrast, the observation, shown in Fig. 6.4(a) and Fig. 6.4(b), that the positive correlation does increase the infection probabilities of large-degree nodes but decreases those of small-degree nodes more when the recovery rate is small, also supports our explanation about how the positive correlation prohibits the spreading when recovery rate is small.

As the recovery rate becomes large thus the prevalence is low, the positive but not negative correlation, between the infection rates and nodal degrees may effectively enhance the infection probabilities of the large degree nodes comparing to the uncorrelated infection rates. When the recovery rate  $\delta$  is large, for example,  $\delta = 5$  for SF networks and  $\delta = 2$  for ER networks in this chapter, the positive correlation leads to the increment of the infection probabilities of large-degree nodes and could further lift the probabilities of their small-degree neighbors which are large in number. The infection probabilities of small-degree nodes are reduced, but the infection probabilities are already low and the small-degree nodes have few neighbor to infect. Hence, the overall infection increases on average. The explanations are supported by Fig. 6.4(a) and Fig. 6.4(b). The infection

<sup>1</sup>We still select  $\alpha = -0.6$  as the example of the negative correlation and  $\alpha = 0.6$  as the comparison for ER network, but we select  $\alpha = -0.2$  as the example of the negative correlation (and  $\alpha = 0.2$  as the comparison) for SF networks since when  $\delta = 5$  and  $\alpha = -0.6$  the epidemic already dies out in the scenario of correlated infection rates.

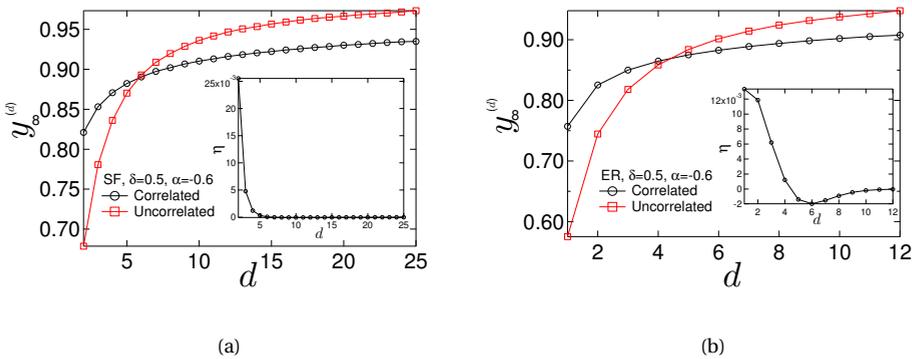


Figure 6.3: In the main figures: the average infection probability  $y_\infty^{(d)}$  of the nodes with degree  $d$  as a function of the degree  $d$  for (a) SF networks and (b) ER networks with the recovery rate  $\delta = 2$  and  $\alpha = -0.6$  for both scenarios of the correlated and uncorrelated infection rates. In the insets:  $\eta$  as a function of  $d$  with the same setting as the main figure.

probabilities of the large-degree nodes increases for different recovery rates when  $\alpha$  is positive, and the increment of the infection probabilities of large-degree nodes is also larger as the recovery rate increases. Though the infection probabilities of the small-degree nodes may decrease when the correlation is negative, the increment of the infection probabilities of the large-degree nodes also lifts the infection probabilities of the small-degree nodes. As a result, the infection probabilities of majority nodes are on average lifted when the recovery rate  $\delta = 5$  and  $\delta = 2$  for SF and ER networks respectively.

For both ER and SF networks, the positive correlation between the nodal degrees and the infection rates tends to enhance the spreading when the recovery rate is small, whereas the negative correlation tends to help when the recovery rate is large. As the recovery rate  $\delta$  increases from 0 and if the absolute value of  $\alpha$  is small, we expect that there is a critical value  $\delta_c$ : when  $\delta < \delta_c$  the negative correlation tends to enhance the spreading, otherwise ( $\delta > \delta_c$ ) the positive correlation is likely to help the spreading. By the comparing Fig. 6.2(c) and Fig. 6.2(f), we can observe that  $\delta_c$  is larger in SF networks than ER networks. This difference is mainly caused by that the prevalence in SF networks tends to be higher than that in ER networks when the recovery rate and the parameter  $\alpha$  are the same and the positive correlation tends to enhance the epidemic spreading when the prevalence is low as we discussed.

### 6.3.2. EXTREME CASES

We then discuss the influence of the correlation between the infection rates and the nodal degrees when the correlation is strong, i.e. the absolute value of  $\alpha$  is large.

When the absolute value of  $\alpha$  is large, the variance of the infection rates is large as well. In this case, most links possess a small infection rate and few have a large infection rate. A large proportion of the links have such a small infection rate that the infection processes driven by the small infection rate will hardly happen. The networks are actually filtered by the small infection rates. The other small proportion of the large infection

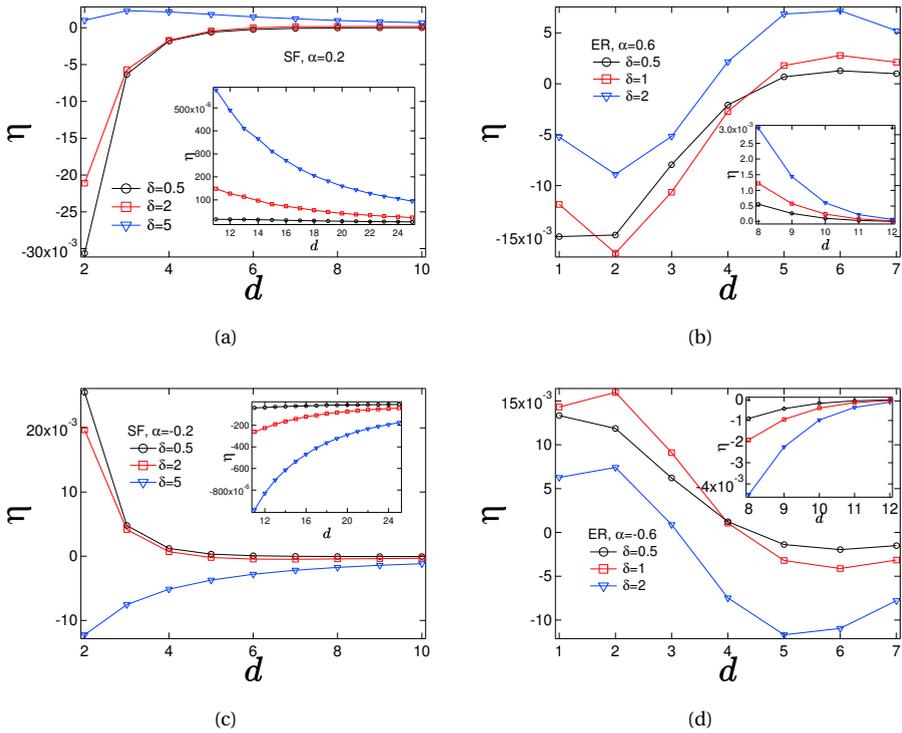


Figure 6.4: The plot of  $\eta$  as a function of  $\alpha$  for (a) SF networks and (b) ER networks with different recovery rates.

rates then mainly determine the overall infection. In the scenario of the uncorrelated infection rates, the few large infection rates are randomly distributed, thus cannot form a connected cluster. Compared to the scenario of the uncorrelated infection rates, the positive correlation ensures that the large infection rates are distributed between the large-degree nodes which are more likely to connect with each other, forming a subgraph. A connected subgraph tends to help the spread of an epidemic. In the other way around, the negative correlation will almost surely stop the epidemic spreading since the few large infection rates distributed between the small-degree nodes could hardly form such a connected cluster. In summary, the positive correlation enhances the epidemic spreading whereas the negative one prohibits when the absolute value of  $\alpha$  is large.

## 6.4. THE WHEEL NETWORK

In this section, we further consider a special topology – the wheel network. We are going to prove that, compared to the uncorrelated heterogeneous infection rate, the negative correlation between the infection rates and the degrees tends to help the epidemic spreading in a large wheel network when the recovery rate is small, whereas the positive correlation tends to contribute to the epidemic spreading when the recovery rate is large.

In a wheel network,  $m$  side nodes compose a ring, i.e. node  $i$  connects with node  $i + 1$  ( $i = 1, 2, 3, \dots, m - 1$ ) and node  $m$  connects with node 1, and all the  $m$  side nodes connect with one central node – node 0. In this section, we consider a large enough wheel network, and without loss of the generality, we still set both the homogeneous infection rate and the average of heterogeneous infection rates to be 1, and tune the recovery rate.

In the scenario of the uncorrelated infection rates, the infection rates are actually i.i.d. In our previous work, we found that, compared to the homogeneous infection rate, the i.i.d. heterogeneous infection rates always reduce the overall infection if the epidemic can spread out. That is to say, the average fraction of infected nodes reaches the maximum when the infection rates are homogeneous, i.e.  $\alpha = 0$  in the scenario of the uncorrelated infection rates. If the average fraction of infected nodes in the scenario of the correlated infection rates is larger than that when the infection rates are homogeneous, then the correlation enhances the epidemic spreading compared to the uncorrelated case.

We first consider the homogeneous infection rate in a wheel network, where the infection rates are homogeneous thus the same for all links. The infection probability  $v_{0\infty}$  of the central node is

$$v_{0\infty} = 1 - \frac{1}{1 + m\tau v_{i\infty}} = 1 - \frac{\delta}{\delta + m v_{i\infty}} \quad (6.3)$$

where  $v_{i\infty}$  is the infection probability of a side node, which is the same for all the side nodes. The infection probability of the side node is

$$v_{i\infty} = 1 - \frac{1}{1 + 2\tau v_{i\infty} + \tau v_{0\infty}} = 1 - \frac{\delta}{\delta + 2v_{i\infty} + v_{0\infty}} \quad (6.4)$$

By solving (6.3) and (6.4), we obtain the positive solution

$$v_{0\infty} = \frac{m - \delta^2 + 2\delta + \frac{\delta((\delta-1)m + 2\delta + \sqrt{(\delta^2 - 2\delta + 9)m^2 - (4\delta^2 + 12\delta)m + 4\delta^2})}{2m}}{\delta + m} \quad (6.5)$$

and

$$v_{i\infty} = \frac{(1 - \delta)m - 2\delta + \sqrt{(\delta^2 - 2\delta + 9)m^2 - (4\delta^2 + 12\delta)m + 4\delta^2}}{4m} \quad (6.6)$$

We consider a large  $m$  and a constant  $\delta$ . In this case,  $v_{0\infty} \approx 1$  and

$$v_{i\infty} \approx \frac{1 - \delta + \sqrt{\delta^2 - 2\delta + 9}}{4} \quad (6.7)$$

The fraction of infected nodes is then

$$y_{\infty} = \frac{mv_{i\infty} + v_{0\infty}}{m + 1} \approx v_{i\infty}$$

Now we consider the wheel network where the infection rates are correlated as defined before in (6.1). There are two kinds of infection rates in a wheel network: 1) the infection rate  $\beta_0$  between the central node and a side node, i.e.  $\beta_0 \sim (3m)^\alpha$ ; 2) the infection rate  $\beta_1$  between a pair of connected side nodes, i.e.  $\beta_1 \sim (3 * 3)^\alpha$ . Since we consider a large  $m$ , we find  $\beta_0 \gg \beta_1$  if  $\alpha > 0$ , whereas  $\beta_0 \ll \beta_1$  if  $\alpha < 0$ . When the correlation between the infection rates and the degree is positive, the infection rates  $\beta_0 \approx 2$  and  $\beta_1 \approx 0$  and the network becomes a star network. In contrast, when the correlation is negative, the infection rates  $\beta_1 \approx 2$  and  $\beta_0 \approx 0$  and the network becomes a ring. By NIMFA, we can compute the average fraction of infected nodes<sup>2</sup> when  $\alpha > 0$

$$y_{\infty} \approx v_{i\infty} = \frac{2}{2 + \delta} \quad (6.8)$$

When  $\alpha < 0$ , the average fraction of infected nodes<sup>3</sup>

$$y_{\infty} = v_{i\infty} = 1 - \frac{\delta}{4} \quad (6.9)$$

The parameter  $\alpha = 0$  in the scenario of the correlated infection rates indicates the homogeneous infection rate which is the same as  $\alpha = 0$  in the scenario of the uncorrelated infection rates. We have shown that the average fraction of infected nodes reaches the maximum when  $\alpha = 0$  in the scenario of the uncorrelated infection rates, so we further compare the average fraction of infected nodes as shown in (6.7) when the infection rates are homogeneous i.e.  $\alpha = 0$  with that as shown in (6.9) when  $\alpha < 0$  or that as shown in (6.8) when  $\alpha > 0$  to explore the influence of the correlation on the epidemic spreading. By comparing (6.7) and (6.9), we find that when the recovery rate  $\delta < 2$ , the average fraction  $y_{\infty}$  of infected nodes is higher if  $\alpha < 0$  than if  $\alpha = 0$ . Hence, the negative correlation between the infection rates and nodal degrees helps the epidemic spreading if the recovery rate is small, i.e.  $\delta < 2$ . Similarly, if we compare (6.7) and (6.8), we find that when the

<sup>2</sup>Equation (6.8) can be similarly derived by solving the NIMFA equations.

<sup>3</sup>Equation (6.9) can be derived by applying the Laurent series.

recovery rate  $\delta > 2$  the average fraction of infected nodes is higher if  $\alpha > 0$  than if  $\alpha = 0$ , and conclude that the positive correlation between the infection rates and nodal degrees contributes to the epidemic spreading if the recovery rate is large. The theoretical results are consistent with our previous conclusions: when the recovery rate is small, the negative correlation tends to help the epidemic spreading, and as the recovery rate increases to be larger than a critical value, i.e. 2 in this case, the positive correlation enhances the spreading.

## 6.5. REAL-WORLD NETWORKS

As mentioned, the interaction frequency  $\beta_{ij}$  between node  $i$  and node  $j$  in a real-world network can be considered as the infection rate between them and it has been found that  $\beta_{ij} \sim (d_i d_j)^\alpha$  in many networks. In this section, we choose two real-world networks as examples to illustrate how their heterogeneous infection rates affect the spread of SIS epidemics on these networks. We compare the average fraction  $y_\infty$  of infected nodes in the metastable state of the two networks in the two scenario: 1) the scenario of correlated infection rates, where each network is equipped with its original heterogeneous infection rates (but normalized so that the average infection rate is 1) as given in the dataset; 2) the scenario of uncorrelated infection rates, where each network is equipped with the infection rates (normalized as well) in the original dataset but the infection rates are shuffled and reassigned to each link. Our objective is to explore the relation between the infection rates and average fraction of infected nodes in these 2 scenarios for both networks to verify our previous findings.

The first network is the airline network (with 3071 nodes and 15358 links) where the nodes are the airports and the infection rate along a link is the number of flights between the two airports. We construct this network and its infection rates from the dataset of openFlights<sup>4</sup>. The other one is the co-author network (with 39577 nodes and 175692 links) where the nodes are the authors of papers, and the infection rate is the collaboration frequency depending on the number of collaborated papers and the number of authors in those papers[87].

As shown in Fig. 6.5(a), the degree distributions of the airline network and co-author network approximately follow a power law with the slope  $\lambda = 1.5$  and 2.5 respectively. Moreover, we plot the average infection rates  $\langle \beta_{ij} \rangle$  as a function of the product of the two nodal degrees  $d_i d_j$  in Fig. 6.5(b) and 6.5(c) for the airline and co-author networks respectively. We find that, roughly  $\beta_{ij} \sim (d_i d_j)^\alpha$  with  $\alpha = 0.14$  and  $\alpha = -0.12$  for the airline and co-author networks respectively.

We first discuss the case when the recovery rate is small, i.e.  $\delta \in [0.5, 8]$  (as shown in Fig. 6.6), and then the case when  $\delta$  is large, i.e.  $\delta \geq 15$  (as shown in Fig. 5.10), since for both networks the recovery rate  $\delta \in [0.5, 8]$  enables the high prevalence of the epidemic and  $\delta \geq 15$  leads to the low prevalence. The average fraction  $y_\infty$  of infected nodes as a function of the recovery rate  $\delta$  for both infection-rate scenarios is shown in in Fig. 6.6 when the recovery rate  $\delta$  is small. We find that the positive correlation ( $\alpha = 0.14$ ) between the infection rates and nodal degrees in the airline network retards the spread of epidemics, whereas the negative correlation ( $\alpha = -0.12$ ) in the co-author network con-

<sup>4</sup><http://openflights.org/data.html>

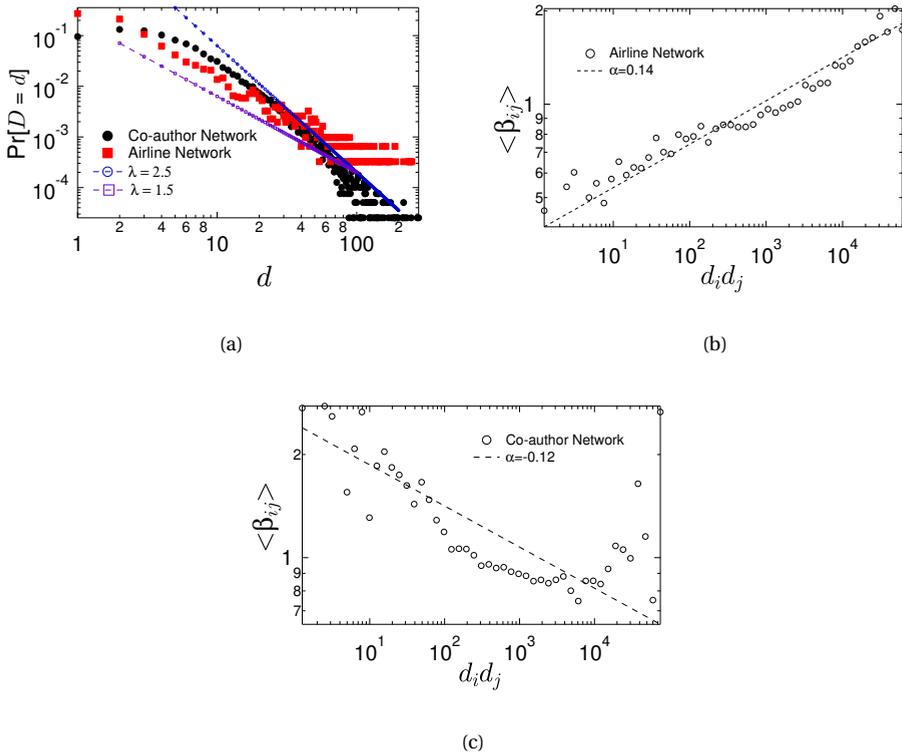


Figure 6.5: (a) The degree distributions of two real-world networks: the airline and co-author networks. (b) The interaction frequency  $\beta_{ij}$  as a function of the nodal degrees  $d_i d_j$  for the airline network. (c) The interaction frequency  $\beta_{ij}$  as a function of the nodal degrees  $d_i d_j$  for the co-author network.

tributes in the other way around. The observations are consistent with our previous conclusions that the positive correlation between the infection rates and nodal degrees probably retards the spread of epidemics whereas a negative correlation may help the spread when the recovery rate is small. The average fraction  $y_\infty$  of infected nodes as a function of the recovery rate  $\delta$  when the recovery rate is large is shown in Fig. 6.7. We find that the positive correlation  $\alpha = 0.14$  in the airline network (Fig. 6.7(a)) helps the epidemic spreading when the recovery rate is larger than 15, and the negative correlation  $\alpha = -0.12$  in the co-author network suppresses the epidemic spreading only when the recovery rate is large, i.e.  $\delta \leq 35$ , that the overall infection is close to 0. The observations also agree with our conclusion that the positive correlation between the infection rates and nodal degrees tends to help the epidemic spreading but not the negative correlation when the recovery rate is large. Hence, the simulation results of the CSIS model on the real-world networks for both the small and large recovery rates agree with our previous conclusions.

## 6.6. CONCLUSION

In this chapter, we study how the correlation between the infection rates and nodal degrees influences the epidemic spreading, compared to the uncorrelated case. By continuous-time simulations of our CSIS model in different infection-rate scenarios and networks with different topology heterogeneities, we find that, when the recovery rate is small, i.e. the prevalence of the epidemic is high, the negative correlation between the nodal degree and the infection rates tends to help the epidemic spreading. However, when the prevalence is high, the positive correlation is more likely to enhance the spreading. The validation on two real-world networks and the proof in the large wheel network agree with our conclusions.

Our results shed light on that how the epidemic spreads in the real-world could be far away from the simple classic models. Not only the heterogeneity of infection rates but also the correlation between the heterogeneity of infection rates and network topologies could be various and complicated. Though we have shown two real-world examples, of which the heterogeneous infection rates obey the assumption in this chapter, i.e., the infection rates are correlated with end-node degrees, there are indeed various types of heterogeneous infection rates. For example, Buono et al. [23] considered a specific distribution of infection rates from the contact frequency between two persons in a real face-to-face experiment. Our work is the first step to study the correlated heterogeneous SIS model in heterogeneous networks.

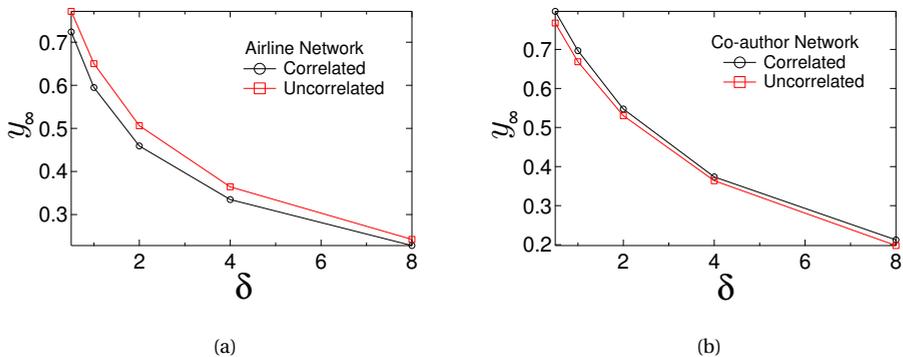


Figure 6.6: The average fraction  $y_\infty$  as a function of the recovery rate  $\delta$  for (a) the airline network and (b) the co-author network in both scenarios of correlated ( $\circ$ ) and uncorrelated ( $\square$ ) infection rates. The recovery rate is in the range  $[0.5, 8]$ .

6

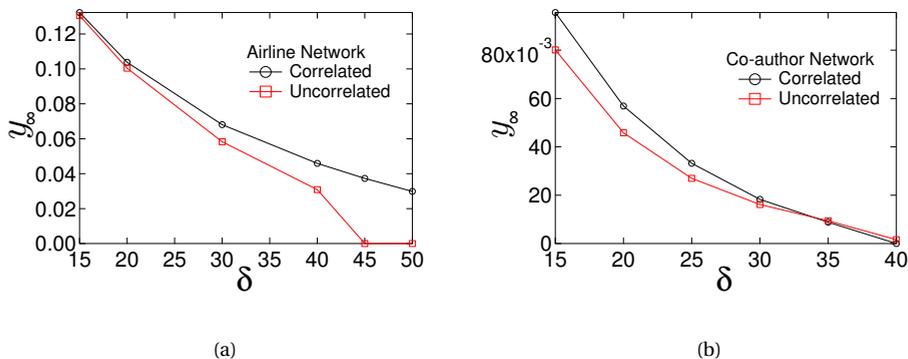


Figure 6.7: The average fraction  $y_\infty$  as a function of the recovery rate  $\delta$  for (a) the airline network and (b) the co-author network in both scenarios of correlated ( $\circ$ ) and uncorrelated ( $\square$ ) infection rates. The recovery rate is large.

# 7

## THE ACCURACY OF MEAN-FIELD APPROXIMATION

*The epidemic spreading over a network has been studied for years by applying the mean-field approach in both homogeneous case, where each node may get infected by an infected neighbor with the same rate, and heterogeneous case, where the infection rates between different pairs of nodes are also different. Researchers have discussed whether the mean-field approaches could accurately describe the epidemic spreading for the homogeneous cases but not for the heterogeneous cases. In this chapter, we explore if and under what conditions the mean-field approach could perform well when the infection rates are heterogeneous. In particular, we employ the Susceptible-Infected-Susceptible (SIS) model and compare the average fraction of infected nodes in the metastable state obtained by the continuous-time simulation and the mean-field approximation. We concentrate on the  $N$ -intertwined Mean Field Approximation (NIMFA), which is an advanced approach considered the underlying network topology. Moreover, for the heterogeneity of the infection rates, we consider not only the independent and identically distributed (i.i.d.) infection rate but also the infection rate correlated with the degree of the two end nodes. We conclude that NIMFA is generally more accurate when the prevalence of the epidemic is higher. Given the same effective infection rate, NIMFA is less accurate when the variance of the i.i.d. infection rate or the correlation between the infection rate and the nodal degree leads to a lower prevalence. Given the same actual prevalence, NIMFA performs better in the cases: 1) when the variance of the i.i.d. infection rates is smaller (while the average is unchanged); 2) when the correlation between the infection rate and the nodal degree is positive. Our work suggests the conditions when the mean-field approach, in particular NIMFA, is more accurate in the approximation of the SIS epidemic with heterogeneous infection rates.*

---

This chapter have been published as: **B. Qu and H. Wang, The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading, Complex Networks 2016.**

## 7.1. INTRODUCTION

By considering the system components or individuals as nodes and the interactions or relations in between nodes as links, networks have been used to describe the biological, social and communication systems. On such networks or complex systems, viral spreading models have been used to describe processes e.g. epidemic spreading and information propagation [33, 73, 79, 92, 128]. The Susceptible-Infected-Susceptible (SIS) model is one of the most studied models. In the SIS model, at any time  $t$ , the state of a node is a Bernoulli random variable, where  $X_i(t) = 0$  represents that node  $i$  is susceptible and  $X_i(t) = 1$  if it is infected. Each infected node infects each of its susceptible neighbors with an infection rate  $\beta$ . The infected node can be recovered with a recovery rate  $\delta$ . Both infection and recovery processes are independent Poisson processes. The ratio  $\tau \triangleq \beta/\delta$  is called effective infection rate, and when  $\tau$  is larger than the epidemic threshold  $\tau_c$ , the epidemic spreads out with a nonzero fraction of infected nodes in the metastable state. The average fraction of infected nodes  $y_\infty$  in the metastable state, ranging in  $[0, 1]$ , indicates how severe the influence of the virus is: the larger the fraction  $y_\infty$  is, the more severely the network is infected.

In this chapter, we concentrate on deriving the average fraction  $y_\infty$  of infected nodes in the metastable state. Although the continuous-time Markov theory can be used to obtain the exact value of  $y_\infty$ , the number of states is too large to be solved in a large network [91]. Hence, the derivation of the average fraction  $y_\infty$  of infected nodes in the metastable state mostly relies on different kinds of mean-field theoretical approaches. The first approach to study the SIS model in complex networks is a degree-based mean-field (DBMF) theory, also called heterogeneous mean-field (HMF) approximation, proposed by Pastor-Satorras et al. [93], which assumes that all nodes with the same degree are statistically equivalent, i.e. the infection probabilities of those nodes are the same. An individual-based mean-field (IBMF) approximations, called the N-Intertwined Mean-Field Approximation (NIMFA), of the SIS model is then introduced [125] with the only assumption that the state of neighboring nodes is statistically independent. A few extensions of the above DBMF and IBMF theories are also developed [16, 36, 48, 84]. NIMFA, taking the network topology into account, turns out to be more precise on different types of networks for the classic SIS model with the homogeneous infection rates [69] while comparing to the DBMF approximation. However, as discussed in [23, 42, 100, 135], the infection rates could be heterogeneous, i.e. the infection rates between different pairs of nodes could also be different. The accuracy of NIMFA, when the infection rates are heterogeneous, has not yet been discussed.

In this chapter, we explore the influence of the heterogeneous infection rates on the precision of NIMFA. In particular, we compare the average fraction  $y_\infty$  of infected nodes as a function of the effective infection rate  $\tau$  computed by NIMFA to that obtained by the continuous-time simulations of the exact SIS model when the infection rates are heterogeneous but the recovery rate is the same for all nodes. In fact, the effective infection rate  $\tau$  refers to the average infection rate divided by the recovery rate in the SIS model with heterogeneous infection rates. We set the average infection rate to 1 and tune the recovery rate  $\delta$  to control the effective infection rate  $\tau$ . We consider both the independent and identically distributed (i.i.d.) and the correlated heterogeneous infection rates in different network topologies. For the case of i.i.d. infection rates, we employ the log-

normal distribution to generate the infection rates as in [100]. In this case, we tune the variance of the infection rates and explore when NIMFA performs better, i.e. the average fraction of infected nodes obtained by NIMFA is closer to that by the continuous-time simulations. For the case of correlated infection rates, we assume that the infection rate  $\beta_{ij}(= \beta_{ji})$  between node  $i$  and  $j$  is correlated with their degrees  $d_i$  and  $d_j$  in the way:

$$\beta_{ij} \sim (d_i d_j)^\alpha \quad (7.1)$$

and  $\alpha$  indicates the strength of the correlation. As discussed in [101], such a correlation between the infection rate and the nodal degree is motivated by the real-world datasets. Moreover, the correlation strength  $\alpha \approx 0.5$  in the network of airports (both in US [11, 81] and China [77]) and  $\alpha \approx 0.8$  in the metabolic network [81]. Given a network, when we generate the heterogeneous infection rates as (6.1), the distribution of infection rates actually changes with the parameter  $\alpha$ , although the average infection rate is kept to be the constant 1. In the case of correlated infection rates, we consider as well the corresponding uncorrelated heterogeneous infection rates scenario, where the correlated infection rates are shuffled and randomly assigned to all the links as a reference scenario, so that we can explore how the correlation between the infection rate and the nodal degree influence the accuracy of NIMFA.

## 7.2. PRELIMINARY

In this section, we introduce the foundation of this chapter, including the i.i.d. heterogeneous infection rates, the correlated heterogeneous infection rates and the continuous-time simulation set up.

### 7.2.1. THE EXTENSION OF NIMFA

The governing equation can be extended to the heterogeneous case:

$$\frac{dv_i(t)}{dt} = -\delta v_i(t) + (1 - v_i(t)) \sum_{j=1}^N \beta_{ij} a_{ij} v_j(t) \quad (7.2)$$

where  $\beta_{ij} = \beta_{ji}$  is the infection rate between node  $i$  and  $j$ . The matrix equation is

$$\left(\frac{1}{\delta} \text{diag}(1 - v_{i\infty}) BA - I\right) V_\infty = 0 \quad (7.3)$$

where  $B$  is the infection rate matrix with the element  $\beta_{ij}$ .

### 7.2.2. THE I.I.D. HETEROGENEOUS INFECTION RATES

In this chapter, we keep the average infection rate to 1 and tune the recovery rate  $\delta$  to control the effective infection rate  $\tau$ . In the case of the i.i.d. heterogeneous infection rates, we aim to explore how the heterogeneous infection rates influence the accuracy of NIMFA when the variance of the infection rate varies. Particularly, we compare the average fraction  $y_\infty$  of infected nodes obtained by NIMFA and the simulations for a given effective infection rate  $\tau$ . In this subsection, we introduce the distribution of the heterogeneous infection rates that will be considered in this work. We choose the infection-rate distribution that is frequently observed in real-world and importantly the variance

is tunable with a fixed mean so that we can systematically explore how the accuracy of NIMFA changes with the broadness of the i.i.d. infection rate.

We consider the log-normal distribution, of which we can keep the mean unchanged and tune the variance in a large range. The log-normal distribution [123]  $B \sim \text{Log-}\mathcal{N}(\beta; \mu, \sigma)$ , of which the probability density function (PDF) is, for  $\beta > 0$

$$f_B(\beta; \mu, \sigma) = \frac{1}{\beta\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln\beta - \mu)^2}{2\sigma^2}\right) \quad (7.4)$$

has a power-law tail for a large range of  $\beta$  provided  $\sigma$  is sufficiently large. The log-normal distribution has as well been widely observed in real-world, where the interaction frequency between nodes is usually considered as the infection rate between those nodes. For example Wang et al. [129] find that by employing the log-normal distributed infection rates, their epidemic model can accurately fit the infection data of 2003 SARS; we also find that the infection rates in an airline network follow the log-normal distribution [100].

In [100], we find that, if the epidemic does not die out, the larger the variance of the i.i.d. infection rate is, the smaller the average fraction  $y_\infty$  of infected nodes is. We will show that this conclusion can actually explain the observation about how the accuracy of NIMFA changes with the variance of the i.i.d. infection rates at a given effective infection rate  $\tau$  in this chapter.

### 7.2.3. THE CORRELATED HETEROGENEOUS INFECTION RATES AND THE RANGE OF $\alpha$

In the case of the correlated heterogeneous infection rates, we build a correlated infection-rate scenario and a reference scenario. In the scenario of correlated infection rates, we assume that  $\beta_{ij} = c(d_i d_j)^\alpha$  where  $c$  is selected so that the average infection rate is 1 and  $\alpha$  indicates the correlation strength. In this case, the infection rate of each link is determined by the given network topology and  $\alpha$ . For the reference scenario, we shuffle the infection rates from all the links as generated in the first scenario and redistribute them randomly to all the links. In this way, we keep the distribution of infection rates but effectively remove the correlation between the infection rates and nodal degrees. For simplicity, we name this reference scenario as the uncorrelated infection-rate scenario. Though the i.i.d. infection rates are also uncorrelated, we can tune the variance of the infection rate in the case of the i.i.d. infection rates while keeping the distribution and the mean of the infection rates. However, in the scenario of uncorrelated infection rates in this chapter, the distribution of the infection rate changes with the parameter  $\alpha$ , hence the variance of the heterogeneous infection rates cannot be systematically tuned.

A positive  $\alpha > 0$  (or negative  $\alpha < 0$ ), suggests a positive (or negative) correlation between the infection rates and nodal degrees. Too large or too small values of  $\alpha$  could not be realistic. For example, [11, 77, 81] suggest that  $\alpha$  is around 0.5 or 0.8 in their datasets. Hence, we select  $\alpha = -0.25, -0.5, -1$  for the negative correlation and  $\alpha = 0.25, 0.5, 1$  for the positive correlation. Different values of  $\alpha$  also offer the possibility to explore how NIMFA performs when the correlation strength is different.

In [101], we find that, comparing to the scenario of uncorrelated infection rate, 1) the positive correlation between the infection rate and the nodal degree tends to increase

(or decrease) the average fraction of infected nodes when the effective infection rate  $\tau$  is small (or large); 2) the negative correlation tends to decrease (or increase) the average fraction  $y_\infty$  of infected nodes when  $\tau$  is small (or large). In this chapter, we aim to understand how the correlation influences the accuracy of NIMFA by comparing the average fraction  $y_\infty$  of infected nodes obtained by NIMFA and the simulations of the exact SIS model. As in the case of the i.i.d. infection rates, we will show that the influence of the correlation between the infection rate and the nodal degree on the average fraction  $y_\infty$  of infected nodes can also be used to partially explain the conclusions in this chapter.

#### 7.2.4. THE SIMULATIONS

We perform the continuous-time simulations of the SIS model on both ER networks and SF networks in this chapter. Given a network topology, a recovery rate  $\delta$ , we carry out 100 iterations. In each iteration, we generate the i.i.d. heterogeneous infection rates following the log-normal distribution or the correlated heterogeneous infection rates as described in (7.1) for the scenario of the correlated infection rates and shuffle them for the scenario of uncorrelated infection rates. Initially, 10% of the nodes are randomly infected. Then the infection and recovery processes of SIS model are simulated until the system reaches the metastable state where the fraction of infected nodes is nonzero and unchanged for a long time if the epidemic spreads out, or the fraction is zero if the epidemic dies out. The average fraction  $y_\infty$  of infected nodes is obtained over 100 iterations.

### 7.3. EFFECT OF THE HETEROGENEOUS INFECTION RATES

In this section, we first explore the accuracy of NIMFA when the heterogeneous infection rates are i.i.d., and particularly how NIMFA performs when the variance  $\text{Var}[B]$  of the infection rate  $B$  varies. Then we explore the influence of the correlated infection rates on NIMFA.

#### 7.3.1. THE I.I.D. INFECTION RATES

We aim to understand the precision of NIMFA under different effective infection rates, different variances of infection rates and different network topologies: we set the average infection rate to 1 and tune the recovery rate  $\delta$  to control the effective infection rate  $\tau$ ; we change the variance of infection rates which follow the log-normal distribution; we consider both ER and SF networks to represent different topologies. For each value of the variance of the infection rate, we obtain the average fraction  $y_\infty$  of infected nodes as a function of the effective infection rate  $\tau$  for NIMFA by numerically solving (7.3) and compare with that by the continuous-time simulations. As shown in Fig. 7.1, no matter what the variance of the infection rate is, the curve of  $y_\infty$  vs.  $\tau$  obtained by NIMFA is close to that obtained by simulations when the actual prevalence of the epidemic is high, i.e. the effective infection rate  $\tau$  is large.

In order to quantify the difference between the two curves obtained by NIMFA and simulations, we define the variable  $\zeta$ :

$$\zeta(\tau) = \frac{|y_{\infty,N}(\tau) - y_{\infty,S}(\tau)|}{y_{\infty,S}(\tau)} \quad (7.5)$$

where  $y_{\infty,N}(\tau) > 0$  and  $y_{\infty,S}(\tau) > 0$  denote the average fraction of infected nodes ob-

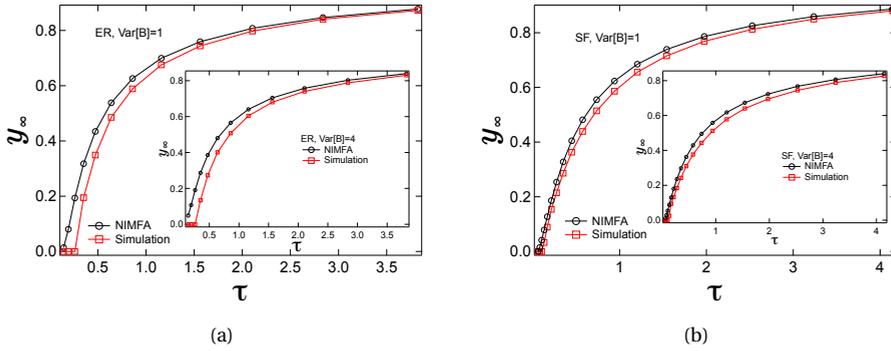


Figure 7.1: The average fraction  $y_\infty$  as a function of the effective infection rate  $\tau$  for (a) ER networks and (b) SF networks. The variances of the infection rates are 1 and 4 in the main figure and the inset respectively.

tained by NIMFA and simulations respectively. The larger the value of  $\zeta(\tau)$  is, the less accurate NIMFA is at the corresponding  $\tau$ .

In Fig. 7.2, the plots of  $\zeta$  vs.  $\tau$  are shown for both ER and SF networks. We find that, for a given effective infection rate  $\tau$ , NIMFA becomes less accurate when the variance of the i.i.d. heterogeneous infection rates increases. This observation can be to a large extent explained by: 1) our finding in Fig. 7.1 that NIMFA is more accurate when the prevalence is higher; 2) that given an effective infection rate  $\tau$  a smaller variance of the i.i.d. infection rates leads to a higher prevalence [100].

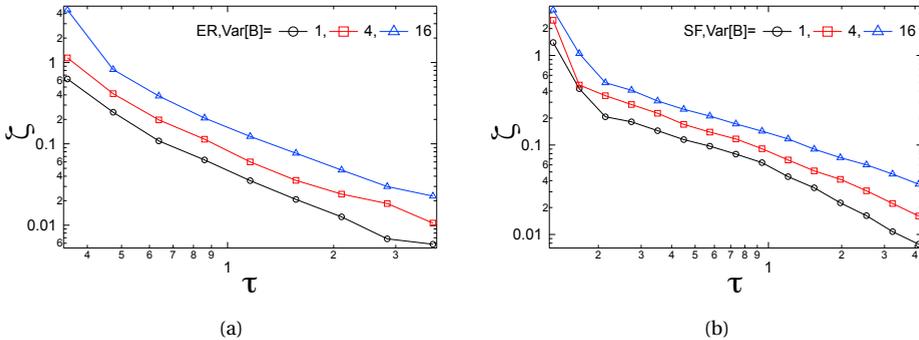


Figure 7.2: The plot of  $\zeta$  as a function of the effective infection rate  $\tau$  for (a) ER networks and (b) SF networks.

We further explore how the variance of the infection rates influences the accuracy of NIMFA if the actual prevalence  $y_{\infty,s}(\tau)$  of epidemic is similar. We plot the variable  $\zeta$  in (7.5) as a function of the actual average fraction of infected nodes obtained by simulations in Fig. 7.3. We find that though it is less evident for ER networks in Fig. 7.3(a), the difference  $\zeta$  in (7.5) is actually larger if the variance of the infection rate is larger as shown in Fig. 7.3(b) for SF networks when the prevalence is the same. Hence, the higher hetero-

generity, i.e. the larger variance, of the i.i.d. infection rates tends to lower down more the accuracy of NIMFA. Overall, we conclude that the prevalence of the epidemic mainly affects the accuracy of NIMFA, i.e. the higher the prevalence is, the more accurate NIMFA tends to be, and given the same prevalence, a larger variance of the i.i.d. infection rates tends to lower down the accuracy of NIMFA.

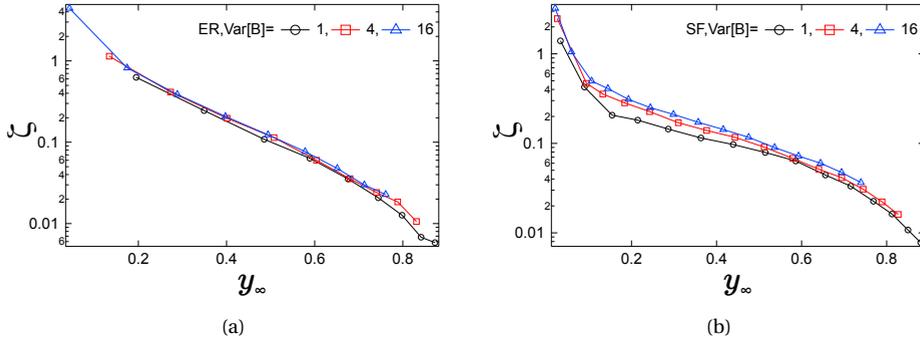


Figure 7.3: The plot of  $\zeta$  as a function of the average fraction  $y_\infty$  obtained by simulations for (a) ER networks and (b) SF networks.

### 7.3.2. THE CORRELATED INFECTION RATE

In this subsection, we aim to understand how the correlation between the infection rate and the nodal degree as shown in (7.1) influences the accuracy of NIMFA. We first employ ER networks as an example and discuss the case when the correlation is positive. Afterwards we explore the influence of the negative correlation.

As mentioned in Section 7.2.3, we build the scenario of uncorrelated infection rates as a reference to study the influence of the correlation between the infection rate and the nodal degree by shuffling the infection rates from all the links as generated in the scenario of correlated infection rates and redistributing them randomly to all the links. As shown in Fig. 7.4(a), we compare the difference  $\zeta$  between NIMFA and simulations in the scenario of uncorrelated and correlated infection rates for both  $\alpha = 0.25$  and  $\alpha = 1$ , and find that  $\zeta$  is smaller in the scenario of correlated infection rates, i.e. NIMFA is more accurate at a given effective infection rate  $\tau$  when the correlation between the infection rate and the nodal degree is positive comparing to the scenario of uncorrelated infection rates. The observations are also consistent with our conclusion that NIMFA is more accurate when the prevalence is higher: the positive correlation tends to increase the average fraction of infected nodes [101], and thus the accuracy of NIMFA, when the effective infection rate  $\tau$  is small; however, when the effective infection rate  $\tau$  is large, though the positive correlate may lower down a bit the average fraction  $y_\infty$  of infected nodes, the prevalence in both scenarios is high, i.e. NIMFA is relatively accurate, and the difference of the accuracy of NIMFA in the two scenarios is not obvious. As the correlation strength  $\alpha$  increases in Fig. 7.4(b), the difference  $\zeta$  decreases at a given  $\tau$ . That is to say, NIMFA tends to be more accurate when the positive correlation becomes stronger.

We further consider the influence of the positive correlation on the accuracy of NIMFA when the prevalence is the same. The plots of the difference  $\zeta$  as a function of the average fraction  $y_\infty$  of infected nodes are shown in Fig. 7.4(c) and Fig. 7.4(d). Given the prevalence of epidemic, the positive correlation is more likely to increase the precision of NIMFA and the stronger the correlation is the more accurate NIMFA is. We observe the same on SF networks which is though not shown here.

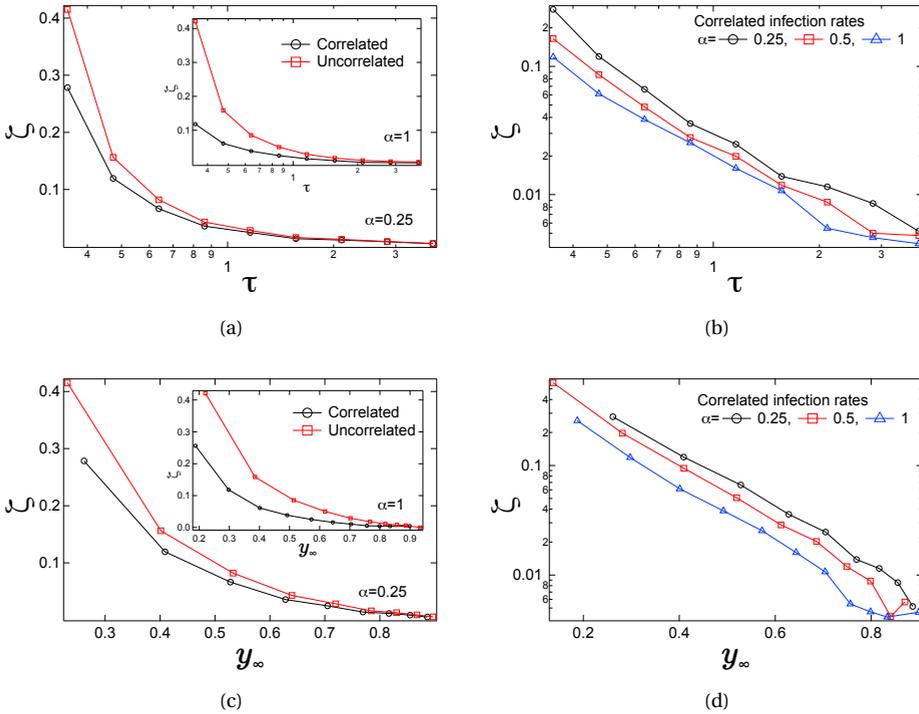


Figure 7.4: (a) The plot of  $\zeta$  as a function of the effective infection rate  $\tau$  in the scenarios of uncorrelated and correlated infection rates for  $\alpha = 0.25$  (the main figure) and  $\alpha = 1$  (the inset). (b) The plot of  $\zeta$  as a function of the effective infection rate  $\tau$  in the scenario of correlated infection rates where different values of  $\alpha$  are considered. (c) The plot of  $\zeta$  as a function of the average fraction  $y_\infty$  of infected nodes obtained by simulations in the scenarios of uncorrelated and correlated infection rates for  $\alpha = 0.25$  (the main figure) and  $\alpha = 1$  (the inset). (d) The plot of  $\zeta$  as a function of the effective infection rate in the scenario of correlated infection rates where different values of  $\alpha$  are considered. All the plots are on ER networks.

Regarding to the influence of the negative correlation between the infection rate and the nodal degree on the accuracy of NIMFA, we compare the variable  $\zeta$  in the scenario of correlated and uncorrelated infection-rate scenario with  $\alpha = -1$  for both ER and SF networks as shown in Fig. 7.5(a). We find that, in general, the negative correlation significantly decreases the accuracy of NIMFA when the effective infection rate  $\tau$  is small but may slightly increase that when  $\tau$  is large. Moreover, NIMFA becomes less accurate when the negative correlation is stronger as shown in Fig. 7.5(b). As mentioned in Section 7.2.3, the negative correlation tends to decrease the prevalence when the effective

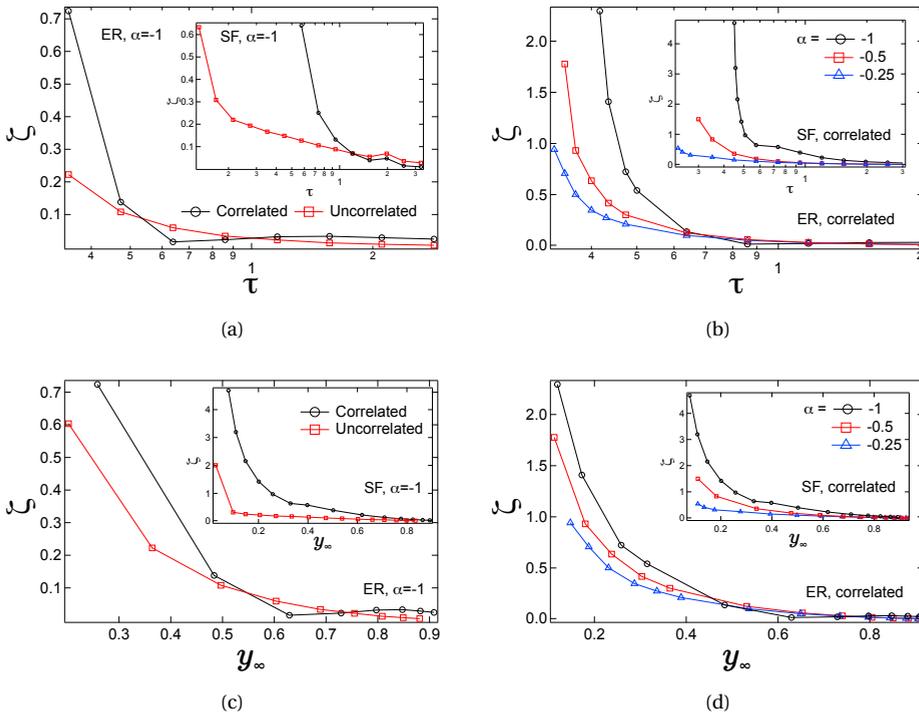


Figure 7.5: (a) The plot of  $\zeta$  as a function of the effective infection rate  $\tau$  in the scenarios of uncorrelated and correlated infection rates for  $\alpha = -1$ . (b) The plot of  $\zeta$  as a function of the effective infection rate  $\tau$  in the scenario of correlated infection rates where different values of  $\alpha$  are considered. (c) The plot of  $\zeta$  as a function of the average fraction  $y_\infty$  of infected nodes obtained by simulations in the scenarios of uncorrelated and correlated infection rates for  $\alpha = -1$ . (d) The plot of  $\zeta$  as a function of the average fraction  $y_\infty$  of infected nodes obtained by simulations in the scenario of correlated infection rates where different values of  $\alpha$  are considered.

infection rate  $\tau$  is small while increase the prevalence when  $\tau$  is large. Hence, the influence of prevalence on the precision of NIMFA could largely explain our observations here.

When the prevalence of epidemic is the same, the influence of the negative correlation on NIMFA's accuracy is shown in Fig. 7.5(c) and Fig. 7.5(d). We find that, in general, 1) NIMFA is less accurate with the negative correlation comparing to the uncorrelated scenario especially when the prevalence is low as shown in Fig. 7.5(c); 2) NIMFA becomes even less accurate if the negative correlation becomes stronger as shown in Fig. 7.5(d).

## 7.4. REAL-WORLD NETWORK

The interaction frequency between two nodes in a real-world network has been considered as the infection rate between the pair of nodes [100]. In this section, we choose the airline network from the real world as an example to illustrate how its heterogeneous

infection rates affect the accuracy of NIMFA of SIS epidemics on the network.

In the airline network, the nodes are the airports, the link between two nodes indicates that there's at least one flight between these two airports, and the infection rate along a link is the number of flights between the two airports. We construct this network and its infection rates from the dataset of openFlights<sup>1</sup>. As shown in [101], the airline network possess roughly a power-law degree distribution. The heterogeneous infection rates from the dataset are normalized by the average so that the average is 1. We compare the difference  $\zeta$  between NIMFA and the simulations of the exact SIS model in three scenarios: 1) the network is equipped with its normalized original heterogeneous infection rates (correlated) as given in the dataset; 2) the network is equipped with the infection rates in the normalized original dataset but randomly shuffled (uncorrelated); 3) the network is equipped with a constant infection rate (homogeneous) which equals to 1. The original heterogeneous infection rate between a pair of nodes are approximately correlated with the degrees of the two nodes as the relationship (7.1), and the parameter  $\alpha \approx 0.14$  indicates a positive correlation [101].

We show the difference  $\zeta$  as a function of the effective infection rate  $\tau$  in Fig. 7.6(a) for the 3 scenarios defined as above. We find that NIMFA is generally more accurate when the effective infection rate  $\tau$  is larger, i.e. the prevalence of epidemic is high. The variable  $\zeta$  is smaller in the scenario of homogeneous infection rates than uncorrelated infection rates with any effective infection rate. This is because the i.i.d. infection rates with a non-zero variance tends to decrease the prevalence, and thus lower down the accuracy of NIMFA at a given effective infection rate  $\tau$ . NIMFA is more accurate with the positive correlation by comparing the difference  $\zeta$  in the scenario of correlated infection rates and uncorrelated infection rates. Furthermore, Fig. 7.6(b) shows that, given the same actual prevalence, i.e. the average fraction  $y_\infty$  of infected nodes obtained by simulations, NIFMA is more accurate: 1) in the homogeneous scenario than in the uncorrelated scenario; 2) in the correlated scenario than in the uncorrelated scenario. All the observations agree with our previous observations and explanations about how the heterogeneous infection rate influences the accuracy of NIMFA in network models.

## 7.5. CONCLUSION

In this chapter, we study how the heterogeneous infection rates affect the accuracy of NIMFA – an advanced mean-field approximation of SIS model that takes the underlying network topology into account. By comparing NIMFA with the continuous-time simulations of the exact SIS model at a give effective infection rate  $\tau$ , we find that the prevalence of epidemic could largely characterize the accuracy of NIMFA which is reflected in two aspects: 1) NIFMA is generally more accurate when the  $\tau$  is larger, i.e. the prevalence of epidemic is higher; 2) when the variance of the i.i.d. infection rates or the correlation between the infection rate and the nodal degree decreases the prevalence at a given  $\tau$ , NIMFA tends to become less accurate as well. Moreover, we also explore the influence of the heterogeneous infection rates on the accuracy of NIMFA at a given prevalence, i.e. when the average fraction  $y_\infty$  of infected nodes obtained by simulations is given. Regarding to the i.i.d. heterogeneous infection rates, the accuracy of NIMFA tends to de-

<sup>1</sup><http://openflights.org/data.html>

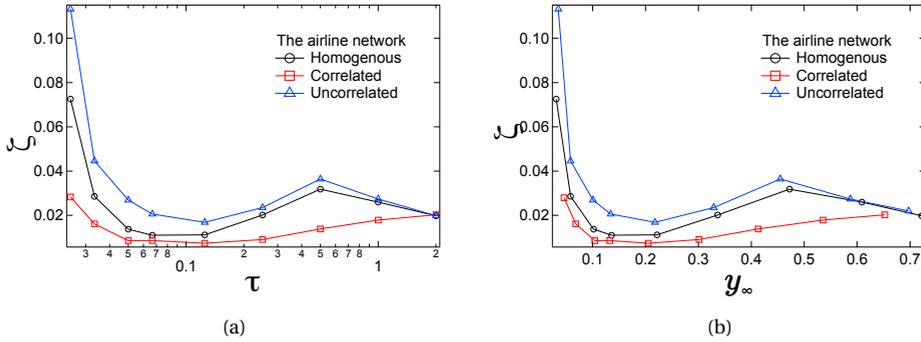


Figure 7.6: The plot of  $\zeta$  as a function of (a) the effective infection rate  $\tau$  and (b) the average fraction  $y_\infty$  of infected nodes obtained by simulations in the airline network with different scenarios of infection rates.

crease as the variance of infection rates increases. In the scenario of correlated infection rates, the positive correlation between the nodal degree and the infection rate is more likely to increase the accuracy of NIMFA whereas the negative correlation tends to lower down the accuracy especially when the effective infection rate is small. Our work sheds light on the conditions when we the mean-field approximation of the SIS model with heterogeneous infection rates is accurate.



# 8

## THE NODAL RANKING OF INFECTION PROBABILITY

*In this chapter, we explore the vulnerability (infection probability) of each node in the metastable state with a given effective infection rate  $\tau$ . Specifically, we investigate the ranking of the nodal vulnerability subject to a susceptible-infected-susceptible epidemic, motivated by the fact that the ranking can be crucial for a network operator to assess which nodes are more vulnerable. Via both theoretical and numerical approaches, we unveil that the ranking of nodal vulnerability tends to change more significantly as  $\tau$  varies when  $\tau$  is smaller or in Barabási-Albert than Erdős-Rényi random graphs.*

---

This chapter have been accepted by Scientific Report as: **B. Qu C. Li, P. Van Mieghem and H. Wang, The nodal infection probability in SIS epidemic spreading.**

## 8.1. INTRODUCTION

The continuous outbreaks of epidemic diseases in a population and viruses in computer networks [3, 45, 46, 60] motivate the study of the epidemic or virus spreading on a network. The Susceptible-infected-susceptible (SIS) epidemic process [15, 28, 33, 69, 73, 111, 121, 128] is widely studied to model the virus spread on a network. In the SIS model, a node is either infected or susceptible at any time  $t$ . Each infected node infects each of its susceptible neighbors with an infection rate  $\beta$ . Each infected node recovers with a recovery rate  $\delta$ . Both infection and recovery processes are independent Poisson processes and the ratio  $\tau = \beta/\delta$  is the effective infection rate. There is an epidemic threshold  $\tau_c$  and a nonzero fraction of nodes is infected in the metastable state when the effective infection rate is above the threshold  $\tau > \tau_c$ . The infection probability  $\nu_{k\infty}(\tau)$  of a node  $k$  in the metastable state at a given effective infection rate  $\tau$  indicates the vulnerability of node  $k$  to the virus, and the average fraction  $y_\infty(\tau)$  of infected nodes reflects the global vulnerability of the network.

Researchers have mainly concentrated on the average fraction  $y_\infty$  of infected nodes in the metastable state to estimate the vulnerability of a network against a certain epidemic or virus. Great effort has been devoted to understand how the network topology influences the vulnerability and the epidemic threshold [72, 91, 121, 135]. The nodal vulnerability or equivalently the infection probability of each node in the metastable state, however, has been seldom studied, except for special cases, i.e. when the effective infection rate is just above the epidemic threshold [123, p. 469]. In this case, it is found that the metastable-state infection probability vector  $V_\infty = [\nu_{1\infty}\nu_{2\infty}\cdots\nu_{N\infty}]^T$ , where  $\nu_{k\infty}$  is the infection probability of node  $k$  in the metastable state) obtained by the N-Intertwined Mean-Field Approximation (NIMFA) of SIS model is proportional to the principal eigenvector  $x_1$  of the adjacency matrix  $A$ . In this chapter, we aim to explore the nodal infection probability in a systematic way, in different network topologies and when the effective infection rate varies. As a starting point, we investigate the ranking of nodal infection probabilities, which crucially informs a network operator regarding to which nodes are more vulnerable or require protection. Interestingly, we find that the ranking of the nodal infection probability changes as the effective infection rate  $\tau$  varies. The observation points out that we cannot find a topological feature of a node to represent the vulnerability of a node to an SIS epidemic, because the rankings in vulnerability of nodes in a network may be different when the effective infection rate  $\tau$  varies, whereas a topological feature of a node remains the same. Our observation explains the finding of Hebert-Dufresne et al. [55] that different nodal features (such as degree, betweenness, etc.) should be used to select the nodes to immunize in different scenarios (based on different infection rates, link densities, etc.), i.e. different nodes should be immunized at different infection rates. In this paper, we explore two questions: (i) which network topology changes the ranking of nodal infection probabilities more significantly and (ii) in which effective infection rate range, the increment of the effective infection rate leads to a more significant change in the ranking for a given network topology?

We first assume that, for an arbitrary pair of nodes, the trajectory  $\nu_{k\infty}(\tau)$  and  $\nu_{m\infty}(\tau)$  cross at most once in any interval  $(\tau_0, \tau_1)$ . We call this assumption the one-crossing assumption and the discussion in Appendix A.2 shows that the assumption is reasonably good. Then The rankings of the vulnerabilities  $\nu_{k\infty}(\tau)$  and  $\nu_{m\infty}(\tau)$  change or equiva-

lently the trajectories  $v_{k\infty}(\tau)$  and  $v_{m\infty}(\tau)$  cross when the effective infection rate  $\tau$  changes from  $\tau_0$  to  $\tau_1$  if  $(v_{k\infty}(\tau_0) - v_{m\infty}(\tau_0))(v_{k\infty}(\tau_1) - v_{m\infty}(\tau_1)) < 0$ . To estimate the maximal change in the ranking of nodal infection probabilities of a network, we consider the total number of crossings between the trajectories of the infection probabilities of all the nodes in a network, when the effective infection rate  $\tau$  changes from just above the epidemic threshold to a large value, above which the ranking remains the same. The total number of crossings is a simple and straightforward measure of the changes in the ranking of nodal infection probabilities. (We also briefly discuss how the correlation of the ranking of nodal infection probabilities changes as the effective infection rate increases in Appendix A.8.) A higher total number of crossings may lead to a more complicated protection policy for a network operator. Given a network, we find a lower bound of the total number of crossings, which can be computed from the topology properties, in particular, from the degree vector and the principal eigenvector of the adjacency matrix. The lower bound is roughly proportional to, thus an accurate indicator of, the total number of crossings for an arbitrary network. Hence, these two topological features could indeed characterize to what extent the ranking of nodal vulnerabilities would change on a network. Since the lower bound is computationally simple, it can be used to compare the total number of crossings for different network topologies. Naturally, this result explains why the total number of crossings tends to be larger in networks with a smaller average degree if the degree distribution is given or with a larger degree variance if the average degree is given. Regarding to question (ii), we analytically derive the number of crossings when the effective infection rate increases with a small value  $\varepsilon$  from a given value  $\tau_0$  if the infection probability vector  $V_\infty(\tau_0)$  at the effective infection rate  $\tau_0$  is also known. This theoretical result, validated by numerical results, explains the observation that the crossings are more likely to occur when the effective infection rate  $\tau$  is smaller.

## 8.2. RESULTS

We first introduce in detail how to count or quantify the changes of the nodal ranking of the infection probability. Afterwards, we investigate the ranking changes (i) in different topologies when the effective infection rate  $\tau$  increases from just above the epidemic threshold to a large value, above which the ranking remains the same, and (ii) when the effective infection rate varies in different ranges.

### 8.2.1. THE COUNTING OF THE NODAL RAKING CHANGES

To explore the changes of the nodal ranking of the infection probability, we investigate in a given network how many crossings, denoted by  $\chi$ , between the trajectory  $v_{k\infty}(\tau)$  and  $v_{m\infty}(\tau)$  for all pairs of nodes can occur in the effective infection rate interval  $(\tau_0, \tau_1)$ , where  $\tau_0 > \tau_c^{(1)}$ . ( $\tau_c^{(1)}$  is the NIMFA threshold epidemic threshold: the epidemic dies out if the effective infection rate  $\tau < \tau_c^{(1)}$ .) In Fig. 8.1, we illustrate the trajectories  $v_{k\infty}(\tau)$  of 10 nodes, randomly selected from a real-world network called Roget ( $N = 994$  nodes, average degree  $E[D] = 7.32$  and detailed in Appendix A.6). For example, the vulnerability of the node corresponding to the red dash line changes dramatically from the medium vulnerable when  $\tau = 0.12$  to the most vulnerable when  $\tau = 0.24$ . Network operators should be alert to such a change of nodal vulnerabilities. The trajectories  $v_{k\infty}(\tau)$  of other groups

of nodes in Roget are shown and discussed in the first section of the supplementary information.

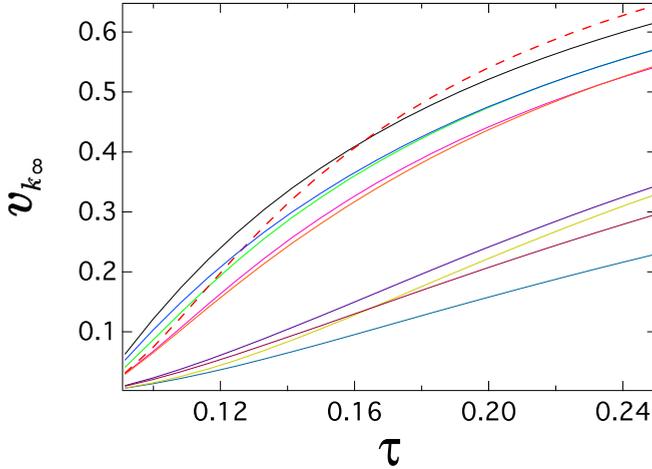


Figure 8.1: The meta-stable infection probability  $v_{k\infty}$  as a function of the effective infection rate  $\tau$  for 10 random nodes in a real-world network called Roget.

For a graph with  $N$  nodes, the maximum number of crossings is  $\frac{N(N-1)}{2}$  with the one-crossing assumption. To count the crossings in the interval  $(\tau_0, \tau_1)$ , we define an  $N \times N$  matrix  $F$  with elements  $f_{ij}$ :

$$f_{ij}(V_\infty(\tau_0), V_\infty(\tau_1)) = (v_{i\infty}(\tau_0) - v_{j\infty}(\tau_0))(v_{i\infty}(\tau_1) - v_{j\infty}(\tau_1))$$

Since  $f_{ii} = 0$ , the matrix  $F$  has a zero diagonal just as the adjacency matrix  $A$ . A negative matrix element  $f_{ij} < 0$  means that there is a crossing between the trajectory  $v_{i\infty}(\tau)$  and  $v_{j\infty}(\tau)$  in the interval  $(\tau_0, \tau_1)$ . The number of crossings in the interval  $(\tau_0, \tau_1)$  of the effective infection rate then equals to

$$\chi(\tau_0, \tau_1) = \sum_{i=1}^N \sum_{j=1}^{i-1} 1_{f_{ij}(V_\infty(\tau_0), V_\infty(\tau_1)) < 0} \quad (8.1)$$

where  $1_{\{x\}}$  is the indicator function:  $1_{\{x\}} = 1$  if the event or condition  $\{x\}$  is true, else  $1_{\{x\}} = 0$ . Specifically, if all nodal degrees are the same in a random graph, the nodal ranking in any intervals of  $\tau$  does not change, since the infection probability of every node [121] equals to the average fraction of infected nodes for any effective infection rate  $\tau$ . In this work, we focus on the nodal infection probability in the meta-stable state, where the initial conditions (such as which and how many nodes are initially infected) has no influence on the meta-state prevalence.

We can compute the SIS metastable viral infection probability  $v_{k\infty}$  of any node  $k$  both by NIMFA [121, 125] and by simulations[69], of the exact SIS continuous-time Markov process. We then further compare the number of crossings  $\chi$  as a function of the increment of the effective infection rate  $\tau$  over different ranges, obtained by NIMFA and the

continuous-time simulations of the SIS model. As shown in Appendix A.7, the number of crossings obtained from NIMFA is relatively close to that from the continuous-time simulations, so we compute the number  $\chi$  of crossings mainly by NIMFA due to its computational efficiency. However, NIMFA may be not accurate when the effective infection rate is close to the epidemic threshold [69]. Hence, the number of crossings obtained by NIMFA and simulations may be quite different from each other when the effective infection rate is close to the epidemic threshold as shown in Appendix A.7.

### 8.2.2. THE TOTAL NUMBER OF CROSSINGS IN DIFFERENT TOPOLOGIES

We explore the total number of crossings in different graph topologies  $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$  when the effective infection rate  $\tau$  changes from just above the epidemic threshold, i.e.  $\tau_c^{(1)} + \varepsilon$ , to a large value  $\tau_u$ , above which the ranking of the nodal infection probability hardly changes. In Appendix A.5, we prove that there exists a value of  $\tau$ , and the ranking of nodal infection probabilities does not change if the effective infection rate is above this value. We derive a lower bound of the total number of crossings and show that the lower bound is actually an accurate indicator of the total number of crossings in different types of graphs.

As shown in Appendix A.3, we derive a lower bound of the total number of crossings in a given graph:

$$\chi_l = \sum_{i=1}^N \sum_{j=1}^{i-1} \mathbf{1}_{f_{ij}(x_1, d) < 0} \leq \chi(\tau_c^{(1)} + \varepsilon, \tau_u) \quad (8.2)$$

where  $x_1$  is the principal eigenvector of the adjacency matrix  $A$ , belonging to the largest eigenvalue  $\lambda_1$  and  $d$  is the degree vector of the given graph.

With the one-crossing assumption, we can compute  $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$  from the infection probability vector  $V_\infty(\tau_c^{(1)} + \varepsilon)$  and  $V_\infty(\tau_u)$ . However, we have to select a proper value of  $\tau_u$  which is large enough and practical. We set the value of  $\tau_u$  as the minimum infection rate such that the average fraction of infected nodes  $y_\infty(\tau_u) \geq 0.9$ , since we find that for most ER, BA random graphs and the aforementioned real-world network, the rankings of the nodal degree and the infection probability are almost the same when the average fraction of infected nodes  $y_\infty \geq 0.9$ . We discuss how we select the value of  $\tau_u$  in Appendix A.5. The scatter plot of the lower bound  $\chi_l$  vs  $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$  is shown in Fig. 8.2 for different graphs including ER random graphs, BA random graphs and six graphs constructed from real-world datasets (as described in Appendix A.6), and the dash line is  $\log \chi_l = \log \chi(\tau_c^{(1)} + \varepsilon, \tau_u) + \log 0.88$ , equivalently

$$\chi_l = 0.88 \chi(\tau_c^{(1)} + \varepsilon, \tau_u) \quad (8.3)$$

We employ the average degree  $E[D] = 8, 10, 12, 14, 16, 18, 20, 40, 60, 80$  for ER random graphs and  $E[D] = 4, 6, 8, 10, 12, 14, 16, 18, 20$  for BA random graphs. Both ER and BA random graphs are with the same size  $N = 1000$ . We confine ourselves to the connected graphs in this work. Hence, we employ the link density  $p = \frac{E[D]}{N-1}$  of ER random graphs, which is larger than the critical link density  $p_c = \frac{\ln N}{N} \approx 0.007$  (equivalently the average degree  $E[D] > 7$ ), to ensure the connectivity. Fig. 8.2 and (8.3) show that the lower bound  $\chi_l$  is indeed always smaller than and approximately proportional to  $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$ . Hence, the lower bound  $\chi_l$  is a computationally simple indication of the total ranking changes of

the metastable state infection probability in a graph. Moreover, we find that for graphs generated by the same random graph model (ER or BA model), a graph with a small average degree tends to have a large number of crossings; given the average degree, a graph with a large degree variance tends to have more crossings. We can understand this observation as follows. The principal eigenvector component of any node  $i$  obeys the eigenvalue equation  $(x_1)_i = \sum_{j=1}^N a_{ij}(x_1)_j$ . The principal eigenvector is positively correlated with the degree vector [68]. Such correlation weakens if the principal eigenvector has a large variance, leading to a large  $\chi_l$ . When the degree variance is large, the variance of the principal eigenvector tends to be as well large, contributing to a large  $\chi_l$ . As more links are added to a network, the network becomes more homogeneous and the variance of the principal eigenvector decreases, resulting in a smaller  $\chi_l$ , or equivalently less crossings.

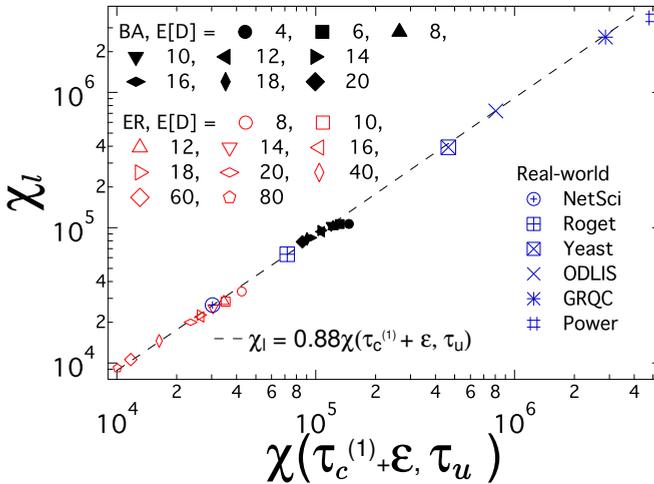


Figure 8.2: The lower bound  $\chi_l$  vs. the total number of crossings  $\chi(\tau_c^{(1)} + \epsilon, \tau_u)$  for ER random graphs ( $\blacktriangle$ ), BA random graphs ( $\blacksquare$ ) and real-world networks ( $\bullet$ )

### 8.2.3. THE NUMBER OF CROSSINGS IN DIFFERENT INTERVALS OF $\tau$

As shown in (8.1), we can compute the number  $\chi(\tau_0, \tau_1)$  of crossings in the given interval  $(\tau_0, \tau_1)$  with the knowledge of the infection probability vectors  $V_\infty(\tau_0)$  and  $V_\infty(\tau_1)$ . Here, we show that we can theoretically derive the number of crossings in a small interval  $(\tau_0, \tau_0 + \Delta\tau)$  with the only knowledge of  $V_\infty(\tau_0)$ . Afterwards, we will validate this theory by numerical results, and illustrate in which ranges of the effective infection rate the number of crossings tends to be larger.

THE CROSSINGS CLOSE TO A GIVEN  $\tau$

For sufficiently small  $\varepsilon = \Delta\tau > 0$ , the Taylor expansion of the steady-state NIMFA infection probability  $v_{k\infty}$  for any node  $k$  is

$$v_{k\infty}(\tau + \varepsilon) = \sum_{m=0}^{\infty} \frac{\varepsilon^m}{m!} \frac{\partial^m v_{k\infty}(\tau)}{\partial \tau^m} = v_{k\infty}(\tau) + \varepsilon \frac{\partial v_{k\infty}(\tau)}{\partial \tau} + \frac{\varepsilon^2}{2} \frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} + O(\varepsilon^3) \quad (8.4)$$

explicit up to order 2. In Appendix A.4, we show the procedure to determine the  $m$ -th order derivative  $v_{i\infty}(\tau)$  with respect to the effective infection rate  $\tau$  for any node  $k$ .

If  $v_{k\infty}(\tau) - v_{m\infty}(\tau) > 0$  and  $\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau} > 0$ , then  $v_{k\infty}(\tau + \varepsilon) - v_{m\infty}(\tau + \varepsilon) > 0$  for sufficiently small  $\varepsilon > 0$  and the ranking at  $\tau + \varepsilon$  and at  $\tau$  is unchanged. On the other hand, if  $v_{k\infty}(\tau + \varepsilon) - v_{m\infty}(\tau + \varepsilon) = 0$ , which implies, for sufficiently small  $\varepsilon > 0$  (so that we can ignore the higher order terms in  $\varepsilon^m$  for  $m > 1$  in (8.4)), that

$$v_{k\infty}(\tau) - v_{m\infty}(\tau) = -\varepsilon \left( \frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau} \right)$$

In other words, given  $v_{k\infty}(\tau)$  of all nodes at  $\tau$ , then there can be a zero or crossing at  $\tau + \varepsilon_{km}$ , where

$$\varepsilon_{km} = -\frac{v_{k\infty}(\tau) - v_{m\infty}(\tau)}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}} \quad (8.5)$$

when  $\varepsilon_{km}$  is small compared to  $\tau$ . This approach is actually known as the Newton-Raphson method and corresponds with the first term in the Lagrange series for the inverse function (see [120] in Page 304). A second order approximation, by ignoring terms of order  $O(\varepsilon^3)$  in (8.4), equating  $v_{k\infty}(\tau + \varepsilon) - v_{m\infty}(\tau + \varepsilon) = 0$  and solving for  $\varepsilon$ , yields

$$\varepsilon_{km} = \frac{-\left(\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}\right) \pm \sqrt{\left(\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}\right)^2 - 2\left(\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}\right)(v_{k\infty}(\tau) - v_{m\infty}(\tau))}}{\left(\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}\right)} \quad (8.6)$$

which is expected to be more accurate, in spite of the higher computational complexity since now also the set of second order derivatives needs to be solved. We rewrite (8.6) as

$$\varepsilon_{km} = -\left(\frac{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}}\right) \left\{ 1 \pm \sqrt{1 - 2\left(\frac{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}\right) \left(\frac{v_{k\infty}(\tau) - v_{m\infty}(\tau)}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}\right)} \right\}$$

Using the generalized binomial expansion  $(1+x)^k = \sum_{k=0}^{\infty} \binom{k}{k} z^k$ , valid for any  $|z| < 1$ , up to first order yields

$$\varepsilon_{km} \approx -\left(\frac{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}}\right) \left\{ 1 \pm \left[ 1 - \left(\frac{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}\right) \left(\frac{v_{k\infty}(\tau) - v_{m\infty}(\tau)}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}\right) \right] \right\}$$

After only retaining the root with the minus sign, we arrive again at (8.5), illustrating that (8.5) is accurate when (8.5) is as small as possible (so that higher order evaluations are not

needed). The discriminant must be positive in order to obtain feasible  $\varepsilon_{km}$ . A positive discriminant is a condition for the existence of crossing in the interval  $(\tau, \tau + \varepsilon)$ . Hence, given an effective infection rate  $\tau_0$  and the corresponding infection probability vector  $V_\infty(\tau_0)$ , there is a crossing close to  $\tau_0$  between the trajectory  $v_{k\infty}(\tau)$  and the trajectory  $v_{m\infty}(\tau)$  at  $\tau_0 + \varepsilon_{km}$  if  $\varepsilon_{km}$  computed by (8.5) is positive and small enough.

#### NUMERICAL RESULTS

In the following of this chapter, we propose to normalize the effective infection rate by the NIMFA epidemic threshold:  $\kappa = \frac{\tau}{\tau_c} \geq 1$ , so that we can compare the number  $\chi$  of crossings in different intervals of  $\kappa$  in the same range  $(1, \kappa_{\max})$  for different network topologies, i.e. different average degrees and different degree distributions. We explore the crossings of the infection probability trajectories when the effective infection rate varies over the range  $(1, \kappa_{\max})$ . We divide the range  $(1, \kappa_{\max})$  into intervals  $(\kappa_{j-1}, \kappa_j)$  where  $j = 1, 2, \dots, n$  is the index and  $\kappa_n = \kappa_{\max}$ .

We aim to explore in which interval of the normalized effective infection rate  $\kappa$  the crossings are more likely to appear. Hence, instead of directly exploring the number of crossings between the trajectory of every node in the whole interval  $(1, \kappa_{\max})$  of the effective infection rate  $\kappa$ , we investigate the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings in (8.1) in each small interval  $(\kappa_{j-1}, \kappa_j)$ . We denote  $\kappa_0 = 1$  (since the effective infection rate below the epidemic threshold corresponds to the all-healthy state),  $\kappa_n = \kappa_{\max}$  and  $\kappa_j = \kappa_0 + j\Delta\kappa$ , where  $\Delta\kappa = (\kappa_{\max} - 1)/n$  is the length of each interval. We will study how the number of crossings changes at different regions of the effective infection rate  $\tau$  or scaled  $\kappa$ . The infection probability  $v_{k\infty}(\kappa)$  at any given value of the normalized effective infection rate  $\kappa$  is computed by solving the NIMFA equation (2.3). On one hand, we can further compute the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings between all node pairs within any interval  $(\kappa_{j-1}, \kappa_j)$  by employing our theoretical results (8.5). On the other hand, we can also numerically compute the number  $\chi(\kappa_{j-1}, \kappa_j)$  by (8.1). We compare the theoretical (8.5) and numerical (8.1) results from the validation point of view. We first compare (8.5) and (8.1) when the normalized effective infection rate  $\kappa$  is not close to 1, i.e. when the effective infection rate  $\tau$  is not close to the epidemic threshold  $\tau_c$ ; specifically, we start from  $\kappa_0 = 2$  and  $\kappa_j = \kappa_0 + j\Delta\kappa$ , where  $\Delta\kappa = 1$ . Figure 8.3 demonstrates that, for both ER and BA graphs, our theoretical result (8.5) agrees well with the numerical result (8.1) except for BA graphs in the interval (2, 3). The less accuracy of our theoretical result for small  $\kappa$  can be explained as follows. Compared to  $\tau_{j-1} = \kappa_{j-1}\tau_c^{(1)}$ , a small value of  $(\kappa_j - \kappa_{j-1})\tau_c^{(1)}$  is required for the accuracy of the theoretical results (8.5), since  $\varepsilon$  in (8.5) is required to be small with respect to the given effective infection rate  $\tau$ . Hence, when  $\kappa_j$  is smaller, a smaller value of  $(\kappa_j - \kappa_{j-1})\tau_c^{(1)}$  is needed for (8.5) to be accurate.

We further plot the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings in the interval  $(\kappa_{j-1}, \kappa_j)$  as a function of  $\kappa_j$ , when the normalized effective infection rate  $\kappa$  is close to 1 and the interval length is reduced to  $\Delta\kappa = 0.1$ . When the interval length, i.e.  $\Delta\kappa$ , is smaller, the theoretical (8.5) results are more consistent with the numerical (8.1) results for BA random graphs in the range of  $\kappa \in (2, 3)$  in Fig. 8.4(b) than in Fig. 8.3(b). For both ER and BA graphs, the two methods agree with each other well when the intervals of  $\kappa$  are small, even when the normalized effective infection rate  $\kappa$  is close to 1 as shown in Fig. 8.4.

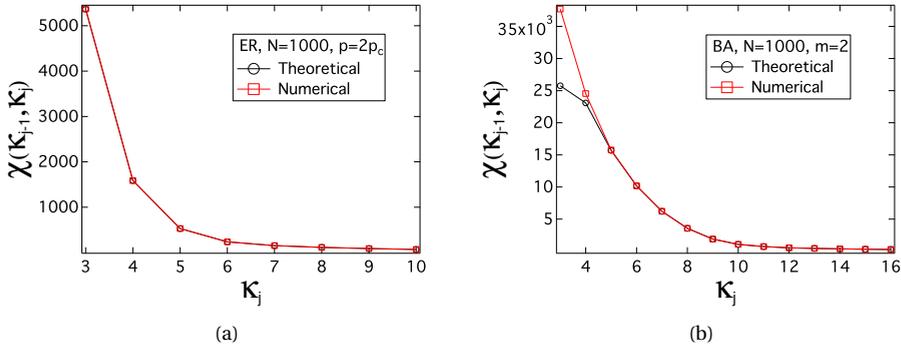


Figure 8.3: The number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings as a function of the normalized effective infection rate  $\kappa_j$ . For ER graphs, we employ the link density  $p = 2p_c$ , thus the average degree  $E[D]=14$ , the size  $N = 1000$  and the NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.0673$ . For BA graphs, we employ the number of newly added links in each step  $m = 2$ , thus the average degree  $E[D] = 4$ , the size  $N = 1000$ , and the NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.0902$ . The results are averaged over 10 realizations.

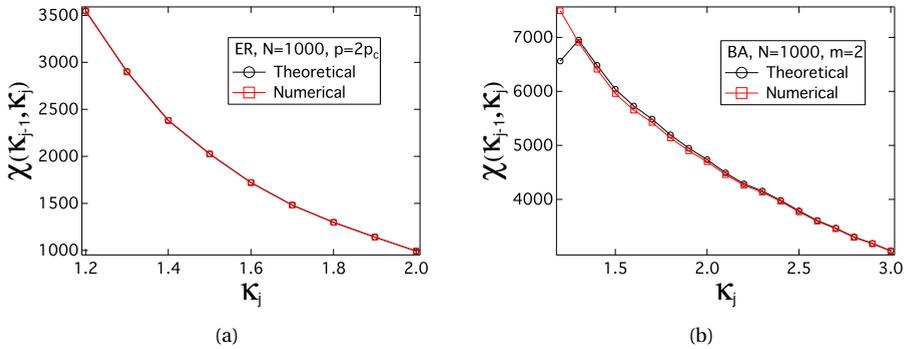


Figure 8.4: The number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings as a function of the normalized effective infection rate  $\kappa_j$ . For ER graphs, we employ the link density  $p = 2p_c$ , thus the average degree  $E[D]=14$ , the size  $N = 1000$  and the NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.0673$ . For BA graphs, we employ the number of newly added links in each step  $m = 2$ , thus the average degree  $E[D] = 4$ , the size  $N = 1000$ , and the NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.0902$ . The results are averaged over 10 realizations.

THE NUMBER OF CROSSINGS IN DIFFERENT  $\tau$  INTERVALS

Both Fig. 8.3 and Fig. 8.4 show that more crossings appear when the effective infection rate is smaller. In this section, we give physical explanations of such observations.

At an effective infection rate  $\tau$  or a normalized effective infection rate  $\kappa$ , Equation (A.1) shows that the comparison of the infection probabilities  $v_{k\infty}(\kappa)$  and  $v_{m\infty}(\kappa)$  is actually equivalent to the comparison of the sum of the infection probabilities of their neighbors, i.e.  $\sum_{j=1}^N a_{kj} v_{j\infty}(\kappa)$  and  $\sum_{j=1}^N a_{mj} v_{j\infty}(\kappa)$ . Without loss of the generality, we assume that the degree  $d_k$  of node  $k$  is larger than the degree  $d_m$  of node  $m$ , i.e.  $d_k > d_m$ . As discussed in Appendix A.3, the infection probability  $v_{k\infty}(\kappa) > v_{m\infty}(\kappa)$  when the effective infection rate is large enough. If there exists a value of the scaled effective infection rate  $\kappa_1$  with which  $\sum_{j=1}^N a_{kj} v_{j\infty}(\kappa_1) < \sum_{j=1}^N a_{mj} v_{j\infty}(\kappa_1)$  while  $d_k > d_m$ , there must be a crossing between  $v_{k\infty}(\kappa)$  and  $v_{m\infty}(\kappa)$  in the interval  $(\kappa_1, \infty)$ . If the infection probability  $v_{j\infty}(\kappa)$  of a node may vary in a larger range with respect to the average infection probability  $\frac{1}{N} \sum_{j=1}^N v_{j\infty}$ , i.e. the average fraction  $y_\infty$  of infected nodes, there may be a higher probability that  $\sum_{j=1}^N a_{kj} v_{j\infty}(\kappa) < \sum_{j=1}^N a_{mj} v_{j\infty}(\kappa)$  and thus more crossings could be expected when the effective infection rate increases from  $\kappa_1$ . Such a hypothesis further motivates us to study the normalized standard deviation of the nodal infection probability:

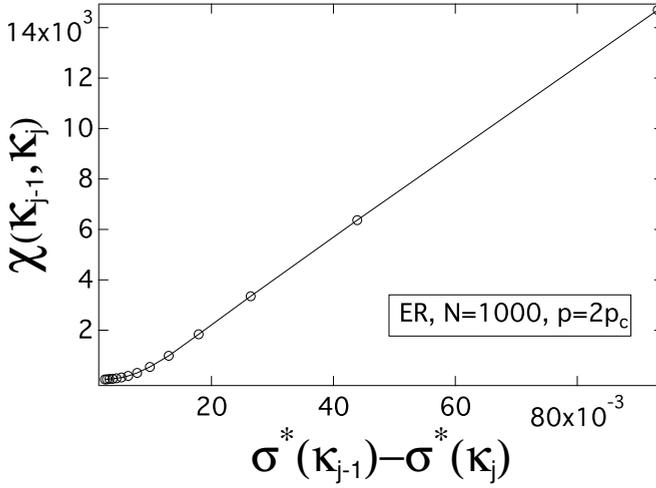
$$\sigma^*(\kappa) = \frac{\sqrt{\sum_{i=1}^N (v_{i\infty}(\kappa) - y_\infty(\kappa))^2 / N}}{y_\infty(\kappa)} \quad (8.7)$$

(where we define  $\sigma^*(\kappa = 1) = \lim_{\kappa \downarrow 1} \sigma^*(\kappa)$ ) and explore whether a larger difference  $|\sigma^*(\kappa_{j-1}) - \sigma^*(\kappa_j)|$  of  $\sigma^*$  would imply more crossings in the interval  $(\kappa_{j-1}, \kappa_j)$ .

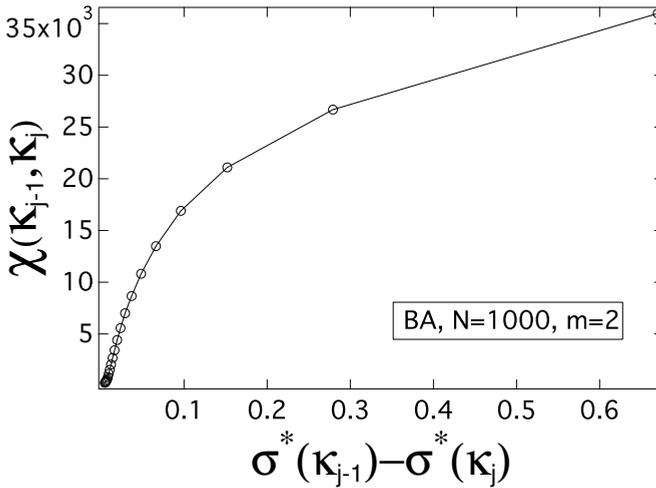
The number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings as a function of the difference  $\sigma^*(\kappa_{j-1}) - \sigma^*(\kappa_j)$  is shown in Fig. 8.5(a) for ER random graphs and in Fig. 8.5(b) for BA random graphs. For both ER and BA random graphs, the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings are positively correlated with the difference  $\sigma^*(\kappa_{j-1}) - \sigma^*(\kappa_j)$  in the interval  $(\kappa_{j-1}, \kappa_j)$ . We observe the same in ER and BA graphs with various average degrees though not shown here. The numerical results support that more crossings tend to appear in an interval where the variable  $\sigma^*$  changes more.

We then further explore how the value of the variable  $\sigma^*(\kappa)$  changes with the normalized effective infection rate  $\kappa$ . We plot the variable  $\sigma^*$  as a function of the normalized effective infection rate  $\kappa$  in Fig. 8.6(a) for ER random graphs and in Fig. 8.6(b) for BA random graphs with  $N = 1000$  and various average degrees, and find that for both types of random graphs the curves can be fitted by a power law function, i.e.  $\sigma^*$  is proportional to  $\kappa^{-\gamma}$ , especially when the average degree is not small. We find that the power law exponent  $\gamma$  of the fitting curves is close to 1 as the average degree  $E[D]$  increases for ER random graphs, and that is always approximately 1 for BA random graphs even though the average degree  $E[D]$  is small. Furthermore, the relationship between the variable  $\sigma^*$  and the normalized effective infection rate  $\kappa$  follows a power law when the normalized effective infection rate  $\kappa$  is much larger as shown in Appendix A.7.

When  $\kappa$  is large, we can theoretically prove the power-law relationship between the variable  $\sigma^*$  and the normalized effective infection rate  $\kappa$ . By (A.9) and assuming a large enough effective infection rate, we obtain  $v_{i\infty}(\tau) = 1 - \frac{1}{\tau d_i} + O(\tau^{-2})$  for node  $i$  and con-

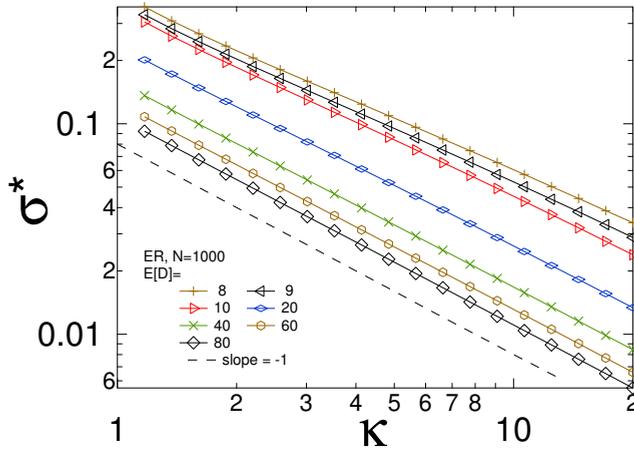


(a)

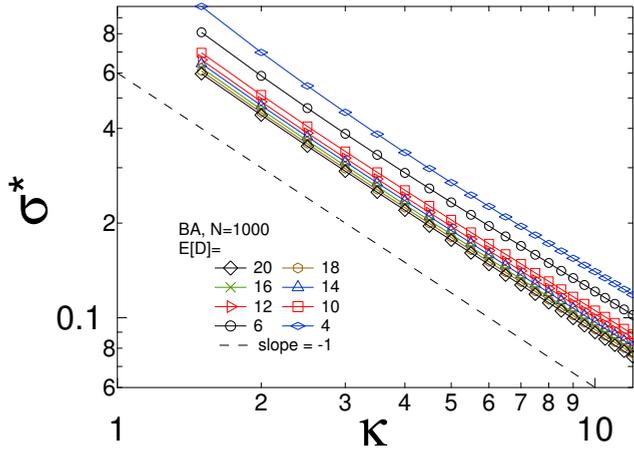


(b)

Figure 8.5: The number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings as a function of the difference  $\sigma^*(\kappa_{j-1}) - \sigma^*(\kappa_j)$  of the normalized standard deviation of the metastable infection probability. For ER graphs, we employ the link density  $p = 2p_c$ , thus the average degree  $E[D] = 14$ , and the size  $N = 1000$  (the NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.0673$ ). For BA graphs, we employ the minimum degree  $m = 2$ , thus the average degree  $E[D] = 4$ , and the size  $N = 1000$  (the NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.0902$ ).



(a)



(b)

Figure 8.6: The normalized standard deviation  $\sigma^*$  of infection probabilities of all nodes as a function of  $\kappa$  in (a) ER and (b) BA random graphs. The dash line is a power-law curve with the exponent  $\gamma = -1$ . The sizes of all random graphs are 1000 and the average degree  $E[D]$  is shown in the figures. The NIMFA epidemic threshold  $\tau_c^{(1)} \approx 0.1097, 0.0993, 0.0902, 0.0476, 0.0244, 0.0164$  and  $0.0124$  for ER random graphs with the average degree  $E[D] = 8, 9, 10, 20, 40, 60$  and  $80$  respectively, and  $\tau_c^{(1)} \approx 0.0902, 0.0698, 0.0479, 0.0416, 0.0368, 0.0329, 0.0300$  and  $0.0274$  for BA random graphs with the average degree  $E[D] = 4, 6, 10, 12, 15, 16, 18$  and  $20$  respectively. The results are averaged over 10 realizations.

sequently  $y_\infty(\tau) = 1 - \frac{1}{\tau}E[\frac{1}{D}] + O(\tau^{-2})$ . By (8.7),

$$\sigma^* = \frac{\sqrt{\text{Var}[\frac{1}{D}]}}{\tau - E[\frac{1}{D}]} + O(\tau^{-2}) \quad (8.8)$$

In a finite graph,  $\text{Var}[\frac{1}{D}]$  and  $E[\frac{1}{D}]$  are finite, hence  $\sigma^*$  is proportional to  $\tau^{-1}$ . The NIMFA epidemic threshold  $\tau_c^{(1)}$  is a constant for a given graph, and with  $\kappa^{-1} = \tau^{-1}\tau_c^{(1)}$ , we obtain that  $\sigma^*$  is proportional to  $\kappa^{-1}$ . Although the power-law relationship between  $\sigma^*$  and  $\kappa$  can be clearly observed in Fig. 8.6, the effective infection rate  $\tau$  corresponding to the variable  $\kappa$  in this figure may be smaller than 1 and the theoretical proof is only valid when the effective infection rate  $\tau \gg 1$ . Our result (8.8) is based on connected graphs, because the terms  $E[\frac{1}{D}]$  and  $\text{Var}[\frac{1}{D}]$  are undefined in unconnected graphs with isolated nodes.

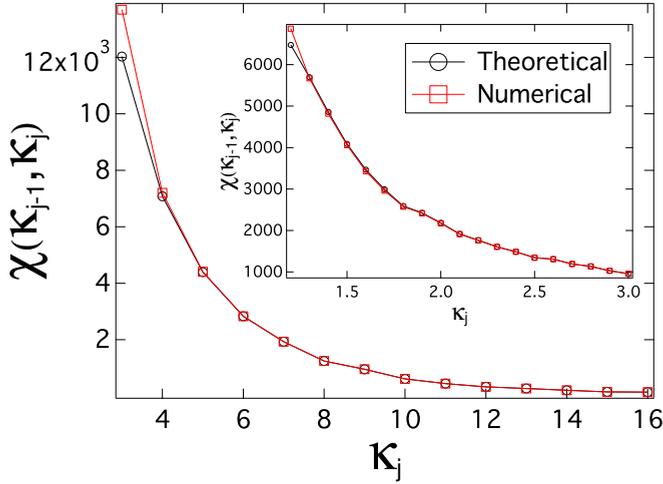
The power-law decay of the variable  $\sigma^*$  with the effective infection rate  $\tau$  explains why there are more crossings when the effective infection rate is smaller by  $\sigma^*$ .

#### VALIDATION ON A REAL-WORLD NETWORK

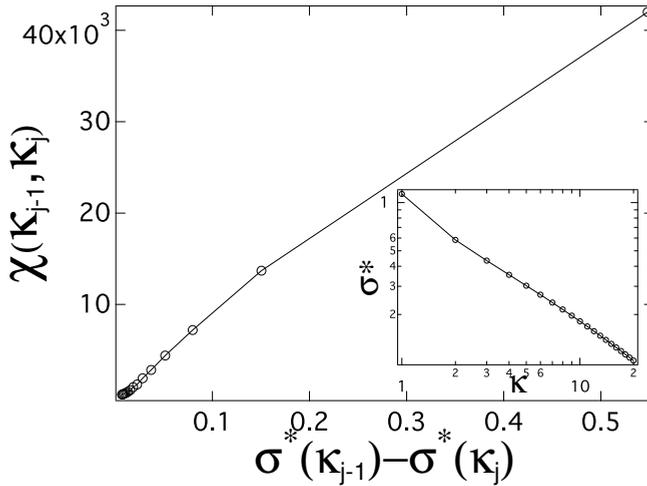
Finally, we validate our previous findings on the real-world network – Roget, detailed in Appendix A.6. As shown in Fig. 8.7(a), the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings at normalized effective infection rate  $\kappa$  interval obtained by theoretical and numerical methods are consistent with each other. The number of crossings decreases fast as  $\kappa$  increases, similar to ER and BA models. The main figure of Fig. 8.7(b) shows that the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings increases with the difference  $\sigma^*(\kappa_{j-1}) - \sigma^*(\kappa_j)$  in the interval  $(\kappa_{j-1}, \kappa_j)$ . In the inset of Fig. 8.7(b), we observe the power-law relationship between the variable  $\sigma^*$  and the normalized effective infection rate  $\kappa$ . All these findings are well in line with previous results on ER and BA random graphs.

### 8.3. DISCUSSION

In the SIS model, the infection probability trajectory  $v_{k\infty}(\tau)$  of node  $k$  and the infection probability trajectory  $v_{m\infty}(\tau)$  of node  $m$  may cross: there exist effective infection rates  $\tau_0$  and  $\tau_1$  and  $(v_{k\infty}(\tau_0) - v_{m\infty}(\tau_0))(v_{k\infty}(\tau_1) - v_{m\infty}(\tau_1)) < 0$ , as the effective infection rate  $\tau$  varies from  $\tau_0$  to  $\tau_1$ . The number of crossings of all node pairs within an interval  $(\tau_0, \tau_1)$  of the effective infection rate measures the change in the ranking of the nodal vulnerabilities when the effective infection rate changes from  $\tau_0$  to  $\tau_1$ . We explore in what types of network topologies and in what ranges of the effective infection rates the crossings are more likely to appear. Theoretically, we find a lower bound  $\chi_l$  in (8.2) of the total number of crossings in a graph. The lower bound  $\chi_l$  only depends on topological features, i.e. the degree vector and principal eigenvector of the adjacency matrix, and is shown to reflect the total number of crossings for a given graph. This allows us to estimate how likely the ranking of the nodal vulnerabilities could change in a given network. Moreover, we analytically predict the crossings close to an effective infection rate  $\tau_0$ , given the infection probabilities of all nodes at the effective infection rate  $\tau_0$ . This theory can be used to estimate the changes of the ranking of the nodal vulnerabilities if the effective infection rate  $\tau$  slightly increases from its current value  $\tau_0$ . We find that more crossings tend to appear when the current effective infection rate, that is subject to changes, is smaller. Our



(a)



(b)

Figure 8.7: (a) The number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings as a function of the normalized effective infection rate  $\kappa_j$ . (b) Main figure: the number  $\chi(\kappa_{j-1}, \kappa_j)$  of crossings as a function of the difference  $\sigma^*(\kappa_{j-1}) - \sigma^*(\kappa_j)$  of the normalized standard deviation of the metastable infection probability; Inset: the normalized standard deviation  $\sigma^*$  of infection probabilities of all nodes as a function of  $\kappa$ .

findings may help network operators to estimate how significant the ranking of nodal vulnerabilities may change for a given change of the effective infection rate on a given network.

This chapter inspires interesting further questions. For example, the change of the node infection probabilities can be quantified beyond the number of crossings, depending on the targeting application; the change in the value of infection probabilities can be further taken into account; can we use nodal infection probabilities to more effectively select the nodes to immunize?



# 9

## REFLECTIONS AND RECOMMENDATIONS

Through the work reported in this thesis we explored the role of the heterogeneity in dynamic processes on complex networks. For the heterogeneity in network topology we considered the case of heterogeneous directed networks. For the heterogeneity in dynamic processes we studied two dynamic process models, the NCO model and the SIS spreading model, and we considered the heterogeneous infection or recovery rates. The main question we addressed is how these types of heterogeneity influence the prevalence of an opinion, epidemic or information in general. In this chapter, the main contributions are summarized and the possible future research directions are proposed.

### 9.1. MAIN CONTRIBUTIONS

In addition to networks with heterogeneous degrees, we considered directed networks as an extra heterogeneity dimension of network topology. We characterized a generic directed network with heterogeneous degrees by the following key features: the percentage of unidirectional links, the in and out degree distribution and the correlation between the indegree and outdegree of a node. Correspondingly, we proposed a system of algorithms to generate directed networks with the given or desirable aforementioned three properties. This allowed us to systematically investigate how each property of a directed network may influence the dynamic process on it.

In Chapter 3, we investigated how each of the three directed network features affects the NCO model, especially the critical threshold, i.e., the minimal initial fraction of population for a given opinion such that this opinion survives (forms a giant cluster) in the steady state. We find that networks with more (less) unidirectional (bidirectional) links and a higher indegree and outdegree correlation tend to have a higher critical threshold. In such networks, an opinion needs a larger initial population in order to survive in the steady state. Chapter 3 sheds light on which types of social interactions/network we could stimulate so that different opinions may co-exist.

In Chapter 4, we investigated the SIS epidemic model on directed networks, where we also applied heterogeneous recovery rates. Our objective was to explore how to reduce the epidemic prevalence via allocating the recovery rates heterogeneously, whereas the total recovery rates (the recovery resources) are limited. We developed a strategy that assigns each node a recovery rate that is dependent on the in- and outdegree of that node. Our strategy evidently outperformed the classic homogeneous allocation of recovery resources in general, especially when the given recovery resources are sufficient. Our work illustrates the potential to more effectively suppress (or to accelerate) the spread of epidemics (information) via heterogeneous allocation of recovery rates.

In Chapter 5 and Chapter 6, we explored how heterogeneous infection rates influence the epidemic spreading. In Chapter 5, we modeled the infection rate along each link as an independently and identically distributed (i.i.d.) random variable. The infection rate between two nodes depends largely on e.g. the contact or collaboration frequency of the node pair, which has been recorded in real-world datasets. Motivated by our observations in real-world datasets, we employed the log-normal and gamma distributions and a newly designed symmetric distribution as the distributions of the infection rate. The average and variance of these infection rate distributions can be controlled, which allowed us to systematically investigate how the average, variance, and possibly higher moments influence the prevalence of an epidemic. We compared the heterogeneous SIS epidemic model with the proposed i.i.d. heterogeneous infection rates and the SIS model with the homogeneous infection rate, when both models have the same average infection rate. We found that the variance (and other higher moments) of the infection rates may evidently affect the epidemic spreading. When the prevalence is relatively high, a larger variance of the infection rates reduces the prevalence more, i.e., heterogeneous infection rates suppress the spread of epidemics. However, the heterogeneous infection rates may facilitate the epidemic spread when the prevalence is relatively low. Our findings point out the opposite effects of the i.i.d. heterogeneous infection rates on epidemic spreading at different prevalence levels.

It is possible that the infection rates in real networks are not i.i.d. We found in real-world datasets that the infection rate of a link may actually be correlated with the degree of the two end nodes. Hence, we took this correlation into our model of the heterogeneous infection rates in Chapter 6. We found that the negative correlation between the infection rate of a link and the degrees of the two end nodes tends to help the epidemic spreading, when the prevalence of the epidemic is relatively high. When the prevalence is low, however, the positive correlation is likely to enhance the spreading. At different prevalence levels, the correlation between infection rates and nodal degrees also affects, in opposite ways, the prevalence of an epidemic.

Our findings unravel the complex influence of heterogeneity on the prevalence of epidemic spreading, ranging from the distribution of the infection rates, the correlation between infection rates and network topology to the heterogeneous degrees. The influence, moreover, tends to differ at different prevalence levels. This understanding may inspire the development of optimization strategies to suppress or facilitate the spread of epidemics/information via incentivizing the formation of desirable network topologies and contact frequencies etc.

From the perspective of methodologies, our studies of the heterogeneous SIS model

on heterogeneous networks are based on continuous-time simulations and mean-field approximations and our results have also been validated using real-world datasets. In Chapter 7, we evaluated the accuracy of the mean-field approximations. In general, mean-field approximations are more accurate when the prevalence of an epidemic is higher. Since the larger variance of the i.i.d. infection rates and the stronger correlation between the infection rate and the nodal degree lead to a lower prevalence at a given effective infection rate, NIMFA is also less accurate in such cases. Our work provides the basic guidelines for researchers regarding to conditions under which mean-field approximations can be applied to study heterogeneous spreading process on heterogeneous networks.

Beyond the heterogeneous infection/recovery rates and network topologies, we explored in Chapter 8 the heterogeneous performance of the nodes, i.e. , the infection probability of each node, subject to SIS epidemic spreading. This is in great contrast to previous work, where the average infection probability, or equivalently, the prevalence has been studied. The infection probability of a node suggests the vulnerability of the node. The heterogeneity of the nodal infection probabilities, especially their ranking can be crucial for a network operator to assess which nodes are more vulnerable. Surprisingly, we found that the ranking of nodal vulnerability may change when the effective infection rate varies. Via both theoretical and numerical approaches, we unveil in which type of networks and which range of the effect infection rate the ranking of the nodal vulnerability is more likely to change. The ranking tends to change more dramatically in sparse networks with heterogeneous degrees and when the effective infection rate is small, thus, likely for realistic epidemics propagating on real-world networks. Our findings imply that as the effective infection rate varies, the ranking of nodal vulnerabilities may change, thus the corresponding network risks and protection strategies may also have to be updated.

## 9.2. FUTURE WORK

Though we have thoroughly studied the influence of various types of heterogeneity on dynamic processes, the heterogeneity in real-world is way more complicated. We would like to raise the following promising but challenging research directions.

**The modeling of heterogeneous dynamic processes.** Take the epidemic spreading process as an example. We modeled the heterogeneous infection rates firstly as i.i.d. random variables that follow distribution as observed in real-world and, furthermore, as functions of nodal degrees. We took into account the correlation between the infection rate of a link and the degrees of the two end nodes. However, the heterogeneous infection rates in real-world could be more complex. For example, the infection rate of a link could be correlated with other topological features of the two end nodes, or even of a local community. Moreover, the infection rates could be dependent on time. For example, an individual that is infected by a disease or possesses a piece of information may have different infection rates when (s)he is just infected and when (s)he has been infected for a long time. Challenging questions are how to model such time-dependent infection rates and how they would influence the epidemic spreading.

**Mean-field approximations of SIS model.** We evaluated the precision of the mean-field approximation of SIS model with heterogeneous infection rates. Moreover, we iden-

tified under which conditions the mean-field approximation is more accurate. However, a foundational explanation is still lacking. Significant effort is needed to improve the mean-field approaches for generic heterogeneous dynamic processes on heterogeneous networks.

**The era of data.** We used a few datasets to motivate our study and verify our results. Various kinds of data at individual level, in contrast to aggregated level, have become available. Examples are user interactions, features and activities in general over time, across physical and online world. Such datasets could further enrich our models, ranging from network topologies, dynamic processes to the interaction between systems.

# BIBLIOGRAPHY

- [1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. Rev. Mod. Phys., 74(1):47, 2002.
- [2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. Nature, 401(6749):130–131, 1999.
- [3] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. Nature, 406(6794):378–382, 2000.
- [4] Franklin Allen and Ana Babus. Networks in finance. 2008.
- [5] Linda JS Allen. Some discrete-time SI, SIR, and SIS epidemic models. Mathematical biosciences, 124(1):83–105, 1994.
- [6] Roy M Anderson and Robert McCredie May. Infectious diseases of humans, volume 1. Oxford university press Oxford, 1991.
- [7] Wen-Jie Bai, Tao Zhou, and Bing-Hong Wang. Immunization of susceptible–infected model on scale-free networks. Physica A, 384(2):656–662, 2007.
- [8] Norman TJ Bailey et al. The mathematical theory of infectious diseases and its applications. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
- [9] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.
- [10] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. Physica A, 281(1):69–77, 2000.
- [11] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11):3747–3752, 2004.
- [12] Daniel Ben-Avraham. Exact solution of the nonconsensus opinion model on the line. Phys. Rev. E, 83(5):050101, 2011.
- [13] Edoardo Beretta and Yasuhiro Takeuchi. Global stability of an sir epidemic model with time delays. Journal of mathematical biology, 33(3):250–260, 1995.
- [14] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. Physics reports, 424(4):175–308, 2006.

- [15] Nino Boccarda and Kyeong Cheong. Critical behaviour of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. J Phys A-Math Gen, 26(15):3707, 1993.
- [16] Marian Boguñá, Claudio Castellano, and Romualdo Pastor-Satorras. Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks. Physical review letters, 111(6):068701, 2013.
- [17] Marián Boguná and Dmitri Krioukov. Navigating ultrasmall worlds in ultrashort time. Phys. Rev. Lett., 102(5):058701, 2009.
- [18] Béla Bollobás. Random graphs. Cambridge University Press, Cambridge, 2001.
- [19] Christian Borghesi, Jean Chiche, and Jean-Pierre Nadal. Between order and disorder: a weak law on recent electoral behavior among urban voters? PloS one, 7(7):e39916, 2012.
- [20] Lidia A Braunstein, Sergey V Buldyrev, Reuven Cohen, Shlomo Havlin, and H Eugene Stanley. Optimal paths in disordered complex networks. Phys. Rev. Lett., 91(16):168701, 2003.
- [21] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3):186–198, 2009.
- [22] Armin Bunde and Shlomo Havlin. Fractals and disordered systems. Springer-Verlag New York, Inc., 1991.
- [23] C Buono, F Vazquez, PA Macri, and LA Braunstein. Slow epidemic extinction in populations with heterogeneous infection rates. Physical Review E, 88(2):022813, 2013.
- [24] Guido Caldarelli, Riccardo Marchetti, and Luciano Pietronero. The fractal properties of internet. EPL (Europhysics Letters), 52(4):386, 2000.
- [25] Geoffrey S Canright and Kenth Engø-Monsen. Spreading on networks: a topographic view. Complexus, 3(1-3):131–146, 2006.
- [26] Benjamin A Carreras, Vickie E Lynch, Ian Dobson, and David E Newman. Critical points and transitions in an electric power transmission model for cascading failure blackouts. Chaos: An interdisciplinary journal of nonlinear science, 12(4):985–994, 2002.
- [27] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. Rev. Mod. Phys., 81:591–646, May 2009.
- [28] E Cator and P Van Mieghem. Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis. Phys. Rev. E, 87(1):012811, 2013.

- [29] Qian Chen, Hyunseok Chang, Ramesh Govindan, and Sugih Jamin. The origin of power laws in internet topologies revisited. In INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, volume 2, pages 608–617. IEEE, 2002.
- [30] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. Physical Review Letters, 85:4626–4628, Nov 2000.
- [31] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient immunization strategies for computer networks and populations. Physical Review Letters, 91(24):247901, 2003.
- [32] Alex R Cook, Wilfred Otten, Glenn Marion, Gavin J Gibson, and Christopher A Gilligan. Estimation of multiple transmission rates for epidemics in heterogeneous populations. Proceedings of the National Academy of Sciences, 104(51):20392–20397, 2007.
- [33] Daryl J Daley, Joe Gani, and Joseph Mark Gani. Epidemic modelling: an introduction, volume 15. Cambridge University Press, 2001.
- [34] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. Adv. Complex Syst., 3(01n04):87–98, 2000.
- [35] Sergei N Dorogovtsev and José FF Mendes. Evolution of networks: From biological nets to the Internet and WWW. Oxford University Press, Oxford, 2003.
- [36] Ken TD Eames and Matt J Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proceedings of the National Academy of Sciences, 99(20):13330–13335, 2002.
- [37] Holger Ebel, Lutz-Ingo Mielsch, and Stefan Bornholdt. Scale-free topology of e-mail networks. Phys. Rev. E, 66(3):035103, 2002.
- [38] Paul Erdős and Alfréd Rényi. On random graphs i. Publ. Math. Debrecen, 6:290–297, 1959.
- [39] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In ACM SIGCOMM Computer Communication Review, volume 29, pages 251–262. ACM, 1999.
- [40] Laura Fratiglioni, Hui-Xin Wang, Kjerstin Ericsson, Margaret Maytan, and Bengt Winblad. Influence of social network on occurrence of dementia: a community-based longitudinal study. The Lancet, 355(9212):1315 – 1319, 2000.
- [41] Linton C Freeman. A set of measures of centrality based on betweenness. Sociometry, pages 35–41, 1977.
- [42] Xinchu Fu, Michael Small, David M Walker, and Haifeng Zhang. Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization. Phys. Rev. E, 77(3):036113, 2008.

- [43] Serge Galam. Minority opinion spreading in random geometry. Eur. Phys. J. B, 25(4):403–406, 2002.
- [44] Serge Galam. Local dynamics vs. social mechanisms: A unifying frame. Europhys. Lett., 70(6):705, 2005.
- [45] Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley. The effect of network topology on the spread of epidemics. In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, volume 2, pages 1455–1466. IEEE, 2005.
- [46] Michele Garetto, Weibo Gong, and Don Towsley. Modeling malware spreading dynamics. In INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, volume 3, pages 1869–1879. IEEE, 2003.
- [47] Or Givan, Nehemia Schwartz, Assaf Cygelberg, and Lewi Stone. Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches. J. Theor. Biol., 288:21–28, 2011.
- [48] James P Gleeson. High-accuracy approximation of binary-state dynamics on networks. Physical Review Letters, 107(6):068701, 2011.
- [49] Nele Goeysvaerts, Niel Hens, Marc Aerts, and Philippe Beutels. Model structure analysis to estimate basic immunological processes and maternal risk for parvovirus b19. Biostatistics, 12(2):283–302, 2011.
- [50] Jesús Gómez-Gardenes, Pablo Echenique, and Yamir Moreno. Immunization of real complex communication networks. The European Physical Journal B, 49(2):259–264, 2006.
- [51] Thilo Gross and Bernd Blasius. Adaptive coevolutionary networks: a review. J. R. Soc., Interface, 5(20):259–271, 2008.
- [52] Daniel Gruhl, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In Proceedings of the 13th international conference on World Wide Web, pages 491–501. ACM, 2004.
- [53] Dongchao Guo, Stojan Trajanovski, Ruud van de Bovenkamp, Huijuan Wang, and Piet Van Mieghem. Epidemic threshold and topological structure of susceptible-infectious-susceptible epidemics in adaptive networks. Physical Review E, 88(4):042802, 2013.
- [54] Fred Halsall and Data Links. Computer networks and open systems. Addison-Wesley Publishers, pages 112–125, 1995.
- [55] Laurent Hébert-Dufresne, Antoine Allard, Jean-Gabriel Young, and Louis J Dubé. Global efficiency of local immunization on complex networks. Scientific Reports, 3, 2013.

- [56] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. J. Artif. Soc. Soc. Simul., 5(3):No. 3, 2002.
- [57] Liang Huang, Kwangho Park, and Ying-Cheng Lai. Information propagation on modular networks. Physical Review E, 73(3):035103, 2006.
- [58] Xuqing Huang, Shuai Shao, Huijuan Wang, Sergey V Buldyrev, H Eugene Stanley, and Shlomo Havlin. The robustness of interdependent clustered networks. EPL (Europhysics Letters), 101(1):18002, 2013.
- [59] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. Nature, 407(6804):651–654, 2000.
- [60] Jeffrey O Kephart and Steve R White. Directed-graph epidemiological models of computer viruses. In Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on, pages 343–359. IEEE, 1991.
- [61] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. Nature Physics, 6(11):888–893, 2010.
- [62] G Korniss. Synchronization in weighted uncorrelated complex networks in a noisy environment: Optimization and connections with transport efficiency. Phys. Rev. E, 75(5):051121, 2007.
- [63] Andrei Korobeinikov. Lyapunov functions and global properties for seir and seis epidemic models. Mathematical Medicine and Biology, 21(2):75–83, 2004.
- [64] Andrei Korobeinikov and Graeme C Wake. Lyapunov functions and global stability for sir, sirs, and sis epidemiological models. Applied Mathematics Letters, 15(8):955–960, 2002.
- [65] PL Krapivsky and S Redner. Dynamics of majority rule in two-state interacting spin systems. Phys. Rev. Lett., 90(23):238701, 2003.
- [66] R Lambiotte and S Redner. Dynamics of non-conservative voters. Europhys. Lett., 82(1):18007, 2008.
- [67] Bibb Latane. The psychology of social impact. Am. Psychol., 36(4):343, 1981.
- [68] C Li, H Wang, W De Haan, CJ Stam, and P Van Mieghem. The correlation of metrics in complex networks with applications in functional brain networks. J Stat Phys, 2011(11):P11018, 2011.
- [69] Cong Li, Ruud van de Bovenkamp, and Piet Van Mieghem. Susceptible-infected-susceptible model: A comparison of n-intertwined and heterogeneous mean-field approximations. Phys. Rev. E, 86(2):026116, 2012.

- [70] Cong Li, Huijuan Wang, and Piet Van Mieghem. Bounds for the spectral radius of a graph when nodes are removed. Linear Algebra and its Applications, 437(1):319–323, 2012.
- [71] Cong Li, Huijuan Wang, and Piet Van Mieghem. Degree and principal eigenvectors in complex networks. In NETWORKING 2012, pages 149–160. Springer, 2012.
- [72] Cong Li, Huijuan Wang, and Piet Van Mieghem. Epidemic threshold in directed networks. Physical Review E, 88(6):062802, 2013.
- [73] Daqing Li, Pengju Qin, Huijuan Wang, Chaoran Liu, and Yinan Jiang. Epidemics on interconnected lattices. EPL (Europhysics Letters), 105(6):68004, 2014.
- [74] Michael Y Li and James S Muldowney. Global stability for the seir model in epidemiology. Mathematical biosciences, 125(2):155–164, 1995.
- [75] Qian Li, Lidia A Braunstein, Shlomo Havlin, and H Eugene Stanley. Strategy of competition between two groups based on an inflexible contrarian opinion model. Phys. Rev. E, 84(6):066101, 2011.
- [76] Qian Li, Lidia A. Braunstein, Huijuan Wang, Jia Shao, H. Eugene Stanley, and Shlomo Havlin. Non-consensus opinion models on complex networks. J. Stat. Phys., 151(1-2):92–112, 2013.
- [77] Wei Li and Xu Cai. Statistical analysis of airport network of china. Physical Review E, 69(4):046106, 2004.
- [78] Thomas M Liggett. Stochastic interacting systems: contact, voter and exclusion processes, volume 324. Springer, New York, 1999.
- [79] Meng Liu, Daqing Li, Pengju Qin, Chaoran Liu, Huijuan Wang, and Feilong Wang. Epidemics in interconnected small-world networks. PloS one, 10(3):e0120701, 2015.
- [80] Eduardo López, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. Anomalous transport in scale-free networks. Phys. Rev. Lett., 94(24):248701, 2005.
- [81] PJ Macdonald, E Almaas, and A-L Barabási. Minimum spanning trees of weighted scale-free networks. EPL (Europhysics Letters), 72(2):308, 2005.
- [82] Nilly Madar, Tomer Kalisky, Reuven Cohen, Daniel ben Avraham, and Shlomo Havlin. Immunization and epidemic dynamics in complex networks. The European Physical Journal B, 38(2):269–276, 2004.
- [83] Antonio Majdandzic, Boris Podobnik, Sergey V Buldyrev, Dror Y Kenett, Shlomo Havlin, and H Eugene Stanley. Spontaneous recovery in dynamical networks. Nature Physics, 10(1):34–38, 2014.
- [84] Angélica S Mata and Silvio C Ferreira. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. EPL (Europhysics Letters), 103(4):48003, 2013.

- [85] Adilson E Motter. Cascade control and defense in complex networks. Phys. Rev. Lett., 93(9):098701, 2004.
- [86] Mark Newman. Networks: an introduction. Oxford University Press, Oxford, 2009.
- [87] Mark EJ Newman. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2):404–409, 2001.
- [88] Mark EJ Newman. Assortative mixing in networks. Physical review letters, 89(20):208701, 2002.
- [89] Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Random graphs with arbitrary degree distributions and their applications. Physical Review E, 64(2):026118, 2001.
- [90] Andrzej Nowak, Jacek Szamrej, and Bibb Latané. From private attitude to public opinion: A dynamic theory of social impact. Psychol. Rev., 97(3):362, 1990.
- [91] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. arXiv preprint arXiv:1408.2701, 2014.
- [92] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. Physical Review E, 63(6):066117, 2001.
- [93] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. Physical Review Letters, 86(14):3200, 2001.
- [94] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. Physical Review E, 65(3):036104, 2002.
- [95] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemics and immunization in scale-free networks. Handbook of graphs and networks: from the genome to the internet, pages 111–130, 2005.
- [96] Victor M Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George Pappas. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. In Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on, pages 7486–7491. IEEE, 2013.
- [97] Victor M Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J Pappas. Optimal resource allocation for network protection against spreading processes. IEEE T. Contr Syst. T., 1(1):99–108, 2014.
- [98] Bo Qu, Alan Hanjalic, and Huijuan Wang. Heterogeneous recovery rates against sis epidemics in directed networks. arXiv preprint arXiv:1408.6959, 2014.
- [99] Bo Qu, Qian Li, Shlomo Havlin, H Eugene Stanley, and Huijuan Wang. Nonconsensus opinion model on directed networks. Physical Review E, 90(5):052811, 2014.

- [100] Bo Qu and Huijuan Wang. SIS epidemic spreading with heterogeneous infection rates. arXiv preprint arXiv:1506.07293, 2015.
- [101] Bo Qu and Huijuan Wang. SIS epidemic spreading with correlated heterogeneous infection rates. arXiv preprint arXiv:1608.07327, 2016.
- [102] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. science, 297(5586):1551–1555, 2002.
- [103] Steven Riley, Christophe Fraser, Christl A Donnelly, Azra C Ghani, Laith J Abu-Raddad, Anthony J Hedley, Gabriel M Leung, Lai-Ming Ho, Tai-Hing Lam, Thuan Q Thach, et al. Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions. Science, 300(5627):1961–1966, 2003.
- [104] Carlos P Roca, Moez Draief, and Dirk Helbing. Percolate or die: Multi-percolation decides the struggle between competing innovations. arXiv:1101.0775, 2011.
- [105] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. IEEE/ACM Transactions on Networking (TON), 21(5):1609–1620, 2013.
- [106] Alejandro D Sánchez, Juan M López, and Miguel A Rodriguez. Nonequilibrium phase transitions in directed small-world networks. Phys. Rev. Lett., 88(4):048701, 2002.
- [107] Arsalan Sattari, Maya Paczuski, and Peter Grassberger. Comment on“dynamic opinion model and invasion percolation”. Phys. Rev. Lett., 109(7):79801, 2012.
- [108] N Schwartz, R Cohen, D Ben-Avraham, A-L Barabási, and S Havlin. Percolation in directed scale-free networks. Phys. Rev. E, 66(1):015104, 2002.
- [109] Frank Schweitzer and Laxmidhar Behera. Nonlinear voter models: the transition from invasion to coexistence. Eur. Phys. J. B, 67(3):301–318, 2009.
- [110] Jia Shao, Shlomo Havlin, and H Eugene Stanley. Dynamic opinion model and invasion percolation. Phys. Rev. Lett., 103(1):018701, 2009.
- [111] Hongjing Shi, Zhisheng Duan, and Guanrong Chen. An SIS model with infective medium on complex networks. Physica A, 387(8):2133–2144, 2008.
- [112] Boris Shulgin, Lewi Stone, and Zvia Agur. Pulse vaccination strategy in the sir epidemic model. Bulletin of mathematical biology, 60(6):1123–1148, 1998.
- [113] Georgos Siganos, Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. Power laws and the as-level internet topology. IEEE/ACM Transactions on Networking (TON), 11(4):514–524, 2003.
- [114] DL Smith, J Dushoff, RW Snow, and SI Hay. The entomological inoculation rate and plasmodium falciparum infection in african children. Nature, 438(7067):492–495, 2005.

- [115] Olaf Sporns, Dante R Chialvo, Marcus Kaiser, and Claus C Hilgetag. Organization, development and function of complex brain networks. Trends in cognitive sciences, 8(9):418–425, 2004.
- [116] Dietrich Stauffer and Ammon Aharony. Introduction to percolation theory. CRC press, Boca Raton, 1994.
- [117] K Subrahmanyam and P Greenfield. Online communication and adolescent relationships. Future of Children, 18(18):119–146, 2008.
- [118] Katarzyna Sznajd-Weron and Józef Sznajd. Opinion evolution in closed community. Int. J. Mod. Phys. C, 11(06):1157–1165, 2000.
- [119] Stojan Trajanovski, Javier Martín-Hernández, Wynand Winterbach, and Piet Van Mieghem. Robustness envelopes of networks. Journal of Complex Networks, 2013.
- [120] Piet Van Mieghem. Graph spectra for complex networks. Cambridge University Press, 2010.
- [121] Piet Van Mieghem. The N-intertwined SIS epidemic network model. Computing, 93(2-4):147–169, 2011.
- [122] Piet Van Mieghem. The viral conductance of a network. Computer Communications, 35(12):1494–1506, 2012.
- [123] Piet Van Mieghem. Performance analysis of communications networks and systems. Cambridge University Press, 2014.
- [124] Piet Van Mieghem and Jasmina Omic. In-homogeneous virus spread in networks. arXiv preprint arXiv:1306.2588, 2013.
- [125] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus spread in networks. IEEE/ACM Transactions on Networking, 17(1):1–14, 2009.
- [126] Piet Van Mieghem, Dragan Stevanović, Fernando Kuipers, Cong Li, Ruud van de Bovenkamp, Daijie Liu, and Huijuan Wang. Decreasing the spectral radius of a graph by link removals. Physical Review E, 84:016101, Jul 2011.
- [127] Piet Van Mieghem, Huijuan Wang, Xin Ge, Siyu Tang, and FA Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. Eur. Phys. J. B, 76(4):643–652, 2010.
- [128] Huijuan Wang, Qian Li, Gregorio Agostino, Shlomo Havlin, H Eugene Stanley, and Piet Van Mieghem. Effect of the interconnected network structure on the epidemic threshold. Physical Review E, 88(2):022801, 2013.
- [129] Wenbin Wang, Ziniu Wu, Chunfeng Wang, and Ruifeng Hu. Modelling the spreading rate of controlled communicable epidemics through an entropy-based thermodynamic model. Sci. Sin.-Phys. Mech. Astron., 56(11):2143, 2013.

- 
- [130] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. nature, 393(6684):440–442, 1998.
- [131] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. Nature, 393(6684):440–442, 1998.
- [132] Martín Weller. The Digital Scholar: How Technology Is Transforming Scholarly Practice. 2011.
- [133] Ramón Xulvi-Brunet and Igor M Sokolov. Changing correlations in networks: assortativity and dissortativity. Acta Physica Polonica B, 36:1431, 2005.
- [134] Rui Yang, Bing-Hong Wang, Jie Ren, Wen-Jie Bai, Zhi-Wen Shi, Wen-Xu Wang, and Tao Zhou. Epidemic spreading on heterogeneous networks with identical infectivity. Physics Letters A, 364(3):189–193, 2007.
- [135] Zimo Yang and Tao Zhou. Epidemic spreading in weighted networks: an edge-based mean-field solution. Physical Review E, 85(5):056106, 2012.
- [136] Yicang Zhou and Hanwu Liu. Stability of periodic solutions for an sis model with pulse vaccination. Mathematical and Computer Modelling, 38(3):299–308, 2003.
- [137] Gerd Zschaler, Gesa A. Böhme, Michael Seißinger, Cristián Huepe, and Thilo Gross. Early fragmentation in the adaptive voter model on directed networks. Phys. Rev. E, 85:046107, Apr 2012.

# ACKNOWLEDGEMENTS

The four years of Ph.D. life is a challenging but amazing journey. It is not only to sharpen one's research skills but to lead one to reflect on the beauty and complexity of research and life. Many times in this journey, I imagine the scene when I write acknowledgments – the final part of my four-year work. However, I just want to do my best to express my sincere gratitude when I am here.

There are so many people to thank for helping me get to where I am. First of all, I would like to thank Alan, Piet and Huijuan. I appreciate the strategic meetings with Alan; I benefit very much from the discussions about research directions and mathematics with Piet; specially, I thank Huijuan for your thoughtful and thorough guidance: thanks for your patience to comment on and revise every paper and for your encouragement whenever I am experiencing grief and loss.

I was so lucky that I can be a member in Network Architectures and Services Group (NAS) as well as Multimedia Computing (MMC) Group. In NAS: many thanks to Cong for countless discussions of research and life, and all laughter and tears with you in those days are my eternal memories; thanks to Ruud for questions of the simulator; many thanks to Niels and Jil for the happy times in the office; thanks to Rob, Fernando, Wendy, Xiangrong, Song, Stojan, Norbert, Ebisa, Javier, Dajie for all kinds of help. In MMC: thanks to Raynor for the kind patience on my questions; thanks to Alessio for tasty Tiramisu and other Italian food; thanks to Xinchao and Yi for the help in my daily life; thanks to Cynthia, Judith, Martha, Christina, Babak, Karthik, Ernestasia, Jeroen and Soude for the fun with you in our group.

The support from a warm Chinese community is also essential to me. I am here to say thanks to Wenjie, Mingxiao, Liang, Qin, Haijin, Shuai, Jicheng, Xin, Jia, Yazhou and others for everything that you do.

Without the love from my family and friends in China and other countries, I do not know where I would be now. Thanks to my parents: I am assured of your greetings and cares on my way to happiness; thanks to Lei: you are my strong defense – whenever I feel tired, you are always there; thanks to Ya: you are like my sister to support me all the time; thanks to Ying, Yi, Qier, Yuan, Duo and many many others for accompanying me in my journey.

No matter how hard I imagined, I never thought I would write all my appreciations down in Hurghada – a peaceful and beautiful seaside town in Egypt. I love the life here so much and feel grateful to everything.

*Bo Qu  
Hurghada, November 2016*



# A

## THE NODAL RANKING OF INFECTION PROBABILITY

### A.1. THE CROSSING BEHAVIOR OF THE TRAJECTORIES $v_{k\infty}$

In this section, we use a real-world network as an example to illustrate the crossing behavior by plotting the infection probability  $v_{k\infty}$  as a function of the effective infection rate  $\tau$  for a small number e.g. 10 nodes. The real-world network is called Roget (detailed in Section Real-world graphs), with 994 nodes and the average degree  $E[D] = 7.32$ . If we plot all values of the infection probability  $v_{k\infty}$  as a function of the effective infection rate  $\tau$  for a network with hundreds of nodes, it would be difficult to tell which two curves actually cross. Hence, we sample 10 nodes, but according to different strategies to illustrate the crossing behavior. In Fig. A.1(a), A.1(b), A.1(c) and A.1(d), 10 nodes are randomly selected from all nodes; in Fig. A.1(e), A.1(f) and A.1(g), 10 nodes are random selected from the nodes with degree  $d = 4, 5$  and  $6$  respectively; Thus, the 10 nodes selected have the same degree in each of these three figures; in Fig. SA.1(h), the top 10 nodes with largest degrees are selected. We find that the crossing of a pair of nodes is indeed significant with respect to the value of their infection probabilities, when the nodes have quite different degrees, as shown in Fig. SA.1(a), SA.1(b) and SA.1(c) where the nodes are selected randomly. The crossing is less significant when the nodes have similar degrees as shown in Fig. SA.1(e), SA.1(f), SA.1(g) and SA.1(h). Since most real-world networks have a heavy tail degree distribution, significant crossing/change in infection probability for pairs of nodes is expected when the infection probability varies.

### A.2. DISCUSSION ABOUT THE ONE-CROSSING ASSUMPTION

We assume that the two trajectories  $v_{k\infty}(\kappa)$  and  $v_{m\infty}(\kappa)$  crosses at most once as the effective infection rate  $\kappa$  changes. Although our theoretical result about the lower bound of the total number of crossings does not depend on this assumption, our method to compute the number of crossings does depend on such an assumption. Hence, we discuss whether the assumption is reasonably good.

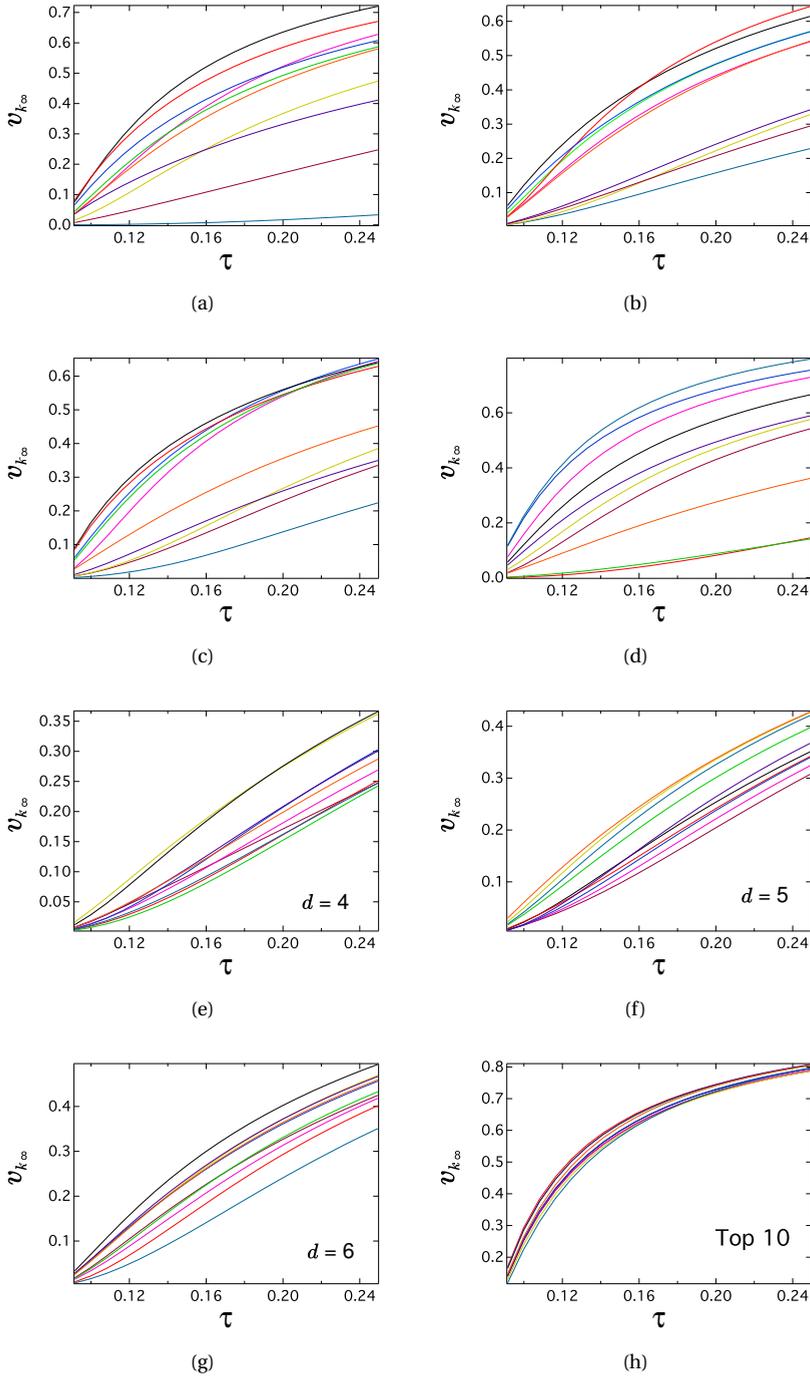


Figure A.1:  $v_{k^{\infty}}$  as a function of  $\tau$  for a real-world network.

Our simulation results so far show that more than one crossing seldom happen. For example, in a real-world network – Roget, only three pairs of nodes have two crossings in their infection probability trajectories in the infection-rate intervals we observed.

When we count numerically the number of crossings between two infection probability curves when  $\kappa$  is changed from 1 to any large value  $\kappa_{max}$ , we divide the interval  $(1, \kappa_{max})$  into a number of  $m$  bins. If  $m = 1$ , we could find maximally one crossing by comparing the infection probability of the two nodes at  $\kappa = 1$  and at  $\kappa = \kappa_{max}$  respectively. As the number of bins increases, we may have the possibility to discover the multiple crossings if they exist. Hence, we explore further whether we observe few node pairs whose infection probability curves cross twice is due to the fact that the bin size we chose is not small enough. Would it be possible that actually two crossings exist within the same bin which would not be observable if we don't split the bin into smaller ones. Hence, we gradually increase the number of bins to explore whether we could find more crossings. As shown in Fig. A.2, we employ ER and BA random graphs with the average degree  $E[D] = 14$  as the examples to show how the number of crossings change when the interval is divided into small ones. We plot the number  $\chi(1, \kappa)$  of crossings as a function of the normalized effective infection rate  $\kappa$ . We do not observe evident increase of the number of crossings (taking all node pairs into account) as the number of bins increases.

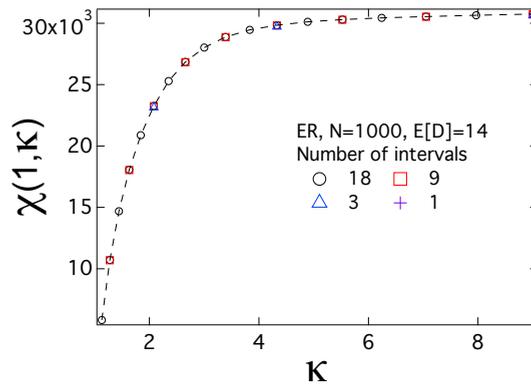
Finally, the bin size should not be too small either. As the bin size becomes small, the change of infection probability for each node when the infection rate is changed from  $\kappa$  to  $\kappa + \epsilon$  is small. In this case, the precision of the numerical solution to compute the infection probability of each node using NIMFA may not be able to distinguish the ranking change of two nodes if their infection probabilities are close. The seemingly two crossings of a node pair may be due to the limited precision of our numerical solution when the bin size is too small.

### A.3. THE DERIVATION OF THE LOWER BOUND $\chi_l$

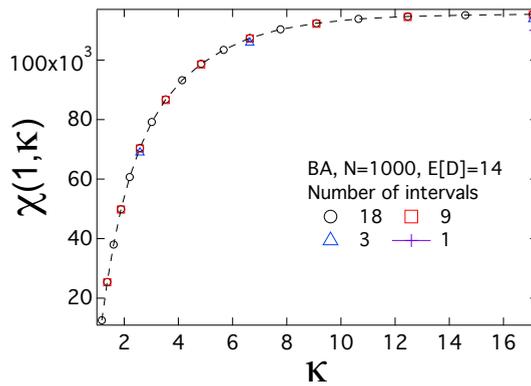
As shown in [123, p. 469] when the effective infection rate  $\tau = \tau_c^{(1)} + \epsilon$  is just above the NIMFA epidemic threshold  $\tau_c^{(1)} = \frac{1}{\lambda_1}$ , the vector  $V_\infty$  with the NIMFA metastable-state infection probabilities is proportional to the principal eigenvector  $x_1$  of the adjacency matrix  $A$ . In particular,  $v_{k\infty} = \epsilon (x_1)_k$ , where  $\epsilon > 0$  and  $(x_1)_k$  is the  $k$ -th component corresponding to node  $k$  of the principal eigenvector  $x_1$  of the adjacency matrix  $A$ , belonging to the largest eigenvalue  $\lambda_1$ . The Perron-Frobenius Theorem [120] states that all vector components of  $x_1$  are non-negative, and even positive if the graph  $G$  is connected. Hence, when the effective infection rate is just above the epidemic threshold, the ranking of the infection probability  $v_{i\infty}(\tau_c^{(1)} + \epsilon)$  is the same as the ranking of the component of the principal eigenvector  $(x_1)_i$ , i.e.  $f_{km}(V_\infty(\tau_c^{(1)} + \epsilon), x_1) = 0$  for any  $k$  and  $m$ .

On the other hand, the NIMFA steady-state infection probability for node  $k$  is given by [125],[123, p. 464] and expressed as

$$v_{k\infty}(\tau) = 1 - \frac{1}{1 + \tau \sum_{j=1}^N a_{kj} v_{j\infty}(\tau)} \quad (\text{A.1})$$



(a)



(b)

Figure A.2: The number  $\chi(1, \kappa)$  of crossings as a function of the normalized effective infection rate  $\kappa$  for (a) ER random graphs and (b) BA random graphs with the same average degree  $E[D] = 14$ .

from which we obtain

$$v_{k\infty}(\tau) - v_{m\infty}(\tau) = \tau (1 - v_{k\infty}(\tau)) (1 - v_{m\infty}(\tau)) \sum_{j=1}^N (a_{kj} - a_{mj}) v_{j\infty}(\tau)$$

The sign of  $v_{k\infty}(\tau) - v_{m\infty}(\tau)$  thus equals to the sign of  $\sum_{j=1}^N (a_{kj} - a_{mj}) v_{j\infty}(\tau)$ . Note that common neighbors of node  $m$  and  $k$  do not play a role in the sign change of  $v_{k\infty}(\tau) - v_{m\infty}(\tau)$ . (The common neighbors of node  $m$  and  $k$  are the set of nodes  $\{j \in \mathcal{N} : a_{mj} = a_{kj}\}$ .) Moreover, if the number of non-common neighbors is 1 (or 0), then there is no change in the sign of  $v_{k\infty}(\tau) - v_{m\infty}(\tau)$  while the effective infection rate  $\tau$  varies. Since the minimum infection probability  $v_{\min}(\tau) > 0$  for  $\tau > \tau_c^{(1)}$  as shown in [123, Lemma 17.4.2 on p. 464], the following bounds apply

$$d_k v_{\min}(\tau) - d_m v_{\max}(\tau) \leq \sum_{j=1}^N (a_{kj} - a_{mj}) v_{j\infty}(\tau) \leq d_k v_{\max}(\tau) - d_m v_{\min}(\tau)$$

where  $v_{\max}(\tau)$  and  $v_{\min}(\tau)$  are the maximum and minimum infection probability respectively and  $d_k$  is the degree of node  $k$ , so that the condition  $v_{k\infty}(\tau) - v_{m\infty}(\tau) > 0$  at  $\tau$  is surely satisfied if  $d_k - d_m \frac{v_{\max}(\tau)}{v_{\min}(\tau)} > 0$ . Using  $v_{\max}(\tau) \leq 1 - \frac{1}{1 + \tau d_{\max}}$  and  $v_{\min}(\tau) \geq 1 - \frac{1}{\tau d_{\min}}$  in [123, p. 464-465], we arrive at the conservative bound for the condition  $v_{k\infty}(\tau) - v_{m\infty}(\tau) > 0$  at  $\tau$ ,

$$d_k > d_m \frac{\tau^2}{\left(\tau - \frac{1}{d_{\min}}\right) \left(\tau + \frac{1}{d_{\max}}\right)}$$

Hence, for large  $\tau$ , the comparison between  $v_{k\infty}(\tau)$  and  $v_{m\infty}(\tau)$  reduces to a comparison in the nodal degree: if  $d_k > d_m$ , then  $v_{k\infty}(\tau) > v_{m\infty}(\tau)$ . This conclusion implies that there exists an effective infection rate  $\tau_u$ , above which the ranking of the metastable-state infection probability is the same as the ranking of the nodal degree, i.e.  $f_{km}(V_\infty(\tau), d) = 0$  for any  $k$  and  $m$  (where  $d$  is the degree vector), if  $\tau \geq \tau_u$ .

The above discussion suggests that the number  $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$  of crossings in the interval  $(\tau_c^{(1)} + \varepsilon, \tau_u)$  is the total number of crossings which a graph can possess. With the one-crossing assumption, we have

$$\chi(\tau_c^{(1)} + \varepsilon, \tau_u) = \sum_{i=1}^N \sum_{j=1}^{i-1} \mathbf{1}_{f_{ij}(V_\infty(\tau_c^{(1)} + \varepsilon), V_\infty(\tau_u)) < 0} \geq \sum_{i=1}^N \sum_{j=1}^{i-1} \mathbf{1}_{f_{ij}(x_1, d) < 0} \quad (\text{A.2})$$

Since only the crossings between two nodes with different degrees are considered in  $\sum_{i=1}^N \sum_{j=1}^{i-1} \mathbf{1}_{f_{ij}(x_1, d) < 0}$ , we obtain a lower bound of the total number  $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$  of crossings. In order to simplify the notation, we denote the lower bound of the total number of crossings by  $\chi_l = \sum_{i=1}^N \sum_{j=1}^{i-1} \mathbf{1}_{f_{ij}(x_1, d) < 0}$ .

#### A.4. DERIVATIVES OF $v_{i\infty}$ WITH RESPECT TO $\tau$

Only for vectors, we use the notation in which a function of a vector is equal to that function applied to the vector components; thus  $f(R) = (f(r_1), f(r_2), \dots, f(r_N))$ . Obviously, this convention does not apply to matrices, where the matrix  $f(A)$  is different than the matrix with elements  $f(a_{ij})$ .

**Theorem 3** Let  $V_\infty$  be the  $N \times 1$  vector with  $k$ -th component  $v_{k\infty}$ , which obeys the NIMFA steady-state equation (A.1). Then, all higher order derivative vectors  $\frac{\partial^m V_\infty}{\partial \tau^m}$  obey the linear equation

$$\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \frac{\partial^m V_\infty}{\partial \tau^m} = R_m \quad (\text{A.3})$$

where  $\mathcal{Q}(q_i) = \text{diag}(q_i) - A$  is the generalized Laplacian and where the right-hand side vector  $R_m$  depends on the previously computed vectors  $(V_\infty, \frac{\partial V_\infty}{\partial \tau}, \dots, \frac{\partial^{m-1} V_\infty}{\partial \tau^{m-1}})$ . In addition, the generalized Laplacian matrix  $\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right)$  and its inverse are positive definite matrices.

**Proof:** Following the approach in [123], the  $i$ -th component of the governing steady-state equation [123, (17.45) on p.466]

$$\frac{1}{\tau} \text{diag} \left( \frac{1}{1-v_{i\infty}} \right) V_\infty = AV_\infty \quad (\text{A.4})$$

written as a generalized Laplacian  $\mathcal{Q}(q_i) = \text{diag}(q_i) - A$ ,

$$\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})} \right) V_\infty = 0$$

is

$$\frac{v_{i\infty}}{1-v_{i\infty}} = \tau \sum_{j=1}^N a_{ij} v_{j\infty}$$

With  $\frac{v_{i\infty}}{1-v_{i\infty}} = \frac{1}{1-v_{i\infty}} - 1$  and  $\sum_{j=1}^N a_{kj} v_{j\infty} = \frac{v_{k\infty}}{\tau(1-v_{k\infty})}$ , differentiation with respect to  $\tau$  yields

$$\frac{1}{(1-v_{k\infty})^2} \frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \tau \sum_{j=1}^N a_{kj} \frac{\partial v_{j\infty}(\tau)}{\partial \tau} = \sum_{j=1}^N a_{kj} v_{j\infty} = \frac{v_{k\infty}}{\tau(1-v_{k\infty})} \quad (\text{A.5})$$

In matrix form, with the definition [123, p. 472] of the generalized Laplacian  $\mathcal{Q}(q_i) = \text{diag}(q_i) - A$ , the vector with the derivatives obeys<sup>1</sup>

$$\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \frac{\partial V_\infty}{\partial \tau} = \frac{1}{\tau^2} \text{diag} \left( \frac{1}{1-v_{i\infty}} \right) V_\infty = \frac{1}{\tau^2} \frac{V_\infty}{1-V_\infty} \quad (\text{A.6})$$

Hence, given the knowledge of the vector  $V_\infty$  at the effective infection rate  $\tau$ , the solution of a linear set returns the components of the vector  $\frac{\partial V_\infty}{\partial \tau}$ .

<sup>1</sup>In [124], we have shown that  $\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right)$  is positive definite (as well as its inverse),

$$\frac{\partial V_\infty}{\partial \tau} = \frac{1}{\tau^2} \mathcal{Q}^{-1} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \text{diag} \left( \frac{1}{1-v_{i\infty}} \right) V_\infty$$

from which

$$\frac{\partial v_{k\infty}(\tau)}{\partial \tau} = \frac{1}{\tau^2} \sum_{j=1}^N \left( \mathcal{Q}^{-1} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \right)_{kj} \frac{v_{j\infty}}{1-v_{j\infty}} \geq 0$$

After a second differentiation with respect to  $\tau$  of (A.5) and some manipulations, we have

$$\frac{1}{\tau(1-v_{k\infty})^2} \frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \sum_{j=1}^N a_{kj} \frac{\partial^2 v_{j\infty}(\tau)}{\partial \tau^2} = \frac{2}{\tau} \sum_{j=1}^N a_{kj} \frac{\partial v_{j\infty}(\tau)}{\partial \tau} - \frac{2}{\tau(1-v_{k\infty})^3} \left( \frac{\partial v_{k\infty}(\tau)}{\partial \tau} \right)^2$$

In matrix form, we obtain

$$\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \frac{\partial^2 V_{\infty}}{\partial \tau^2} = \frac{2}{\tau} A \frac{\partial V_{\infty}}{\partial \tau} - \frac{2}{\tau} \text{diag} \left( \frac{1}{(1-v_{k\infty})^3} \right) \left( \frac{\partial V_{k\infty}(\tau)}{\partial \tau} \right)^2$$

We can avoid the matrix computation  $A \frac{\partial V_{\infty}}{\partial \tau}$ , because (A.6) supplies us with

$$\begin{aligned} A \frac{\partial V_{\infty}}{\partial \tau} &= \text{diag} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \frac{\partial V_{\infty}}{\partial \tau} - \frac{1}{\tau^2} \text{diag} \left( \frac{1}{1-v_{i\infty}} \right) V_{\infty} \\ &= \frac{\frac{\partial V_{\infty}}{\partial \tau}}{\tau(1-V_{\infty})^2} - \frac{1}{\tau^2} \frac{V_{\infty}}{1-V_{\infty}} \end{aligned}$$

while the NIMFA matrix equation (A.4) shows that

$$A \frac{\partial V_{\infty}}{\partial \tau} = \frac{d}{d\tau} \frac{1}{\tau} \frac{V_{\infty}}{1-V_{\infty}} = \frac{d}{d\tau} \frac{1}{\tau} \left( \frac{1}{1-V_{\infty}} - u \right)$$

where  $u = (1, 1, \dots, 1)$  is the all-one vector. Hence,

$$\mathcal{Q} \left( \frac{1}{\tau(1-v_{i\infty})^2} \right) \frac{\partial^2 V_{\infty}}{\partial \tau^2} = R_2 \quad (\text{A.7})$$

with

$$R_2 = \frac{2}{\tau^2} \text{diag} \left( \frac{1}{(1-v_{i\infty})^2} \right) \frac{\partial V_{\infty}}{\partial \tau} - \frac{2}{\tau} \text{diag} \left( \frac{1}{(1-v_{k\infty})^3} \right) \left( \frac{\partial V_{\infty}(\tau)}{\partial \tau} \right)^2 - \frac{2}{\tau^3} \text{diag} \left( \frac{1}{1-v_{i\infty}} \right) V_{\infty}$$

or

$$R_2 = \frac{2}{\tau} \left\{ \frac{d}{d\tau} \frac{1}{\tau} \left( \frac{1}{1-V_{\infty}} - u \right) - \frac{\left( \frac{\partial V_{\infty}(\tau)}{\partial \tau} \right)^2}{(1-V_{\infty})^3} \right\}$$

which is a same matrix equation as in (A.6), but a different right-hand side vector, which can only be determined, after solving (A.6).

A next differentiation with respect to  $\tau$  of

$$\frac{1}{(1-v_{k\infty})^2} \frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \tau \sum_{j=1}^N a_{kj} \frac{\partial^2 v_{j\infty}(\tau)}{\partial \tau^2} = 2 \sum_{j=1}^N a_{kj} \frac{\partial v_{j\infty}(\tau)}{\partial \tau} - \frac{2}{(1-v_{k\infty})^3} \left( \frac{\partial v_{k\infty}(\tau)}{\partial \tau} \right)^2$$

shows that the left-hand side  $L$  and the right-hand side  $R$  derivatives are

$$\begin{aligned} L &= \frac{1}{(1-v_{k\infty})^2} \frac{\partial^3 v_{k\infty}(\tau)}{\partial \tau^3} + \frac{2}{(1-v_{k\infty})^3} \frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} \frac{\partial v_{j\infty}(\tau)}{\partial \tau} - \sum_{j=1}^N a_{kj} \frac{\partial^2 v_{j\infty}(\tau)}{\partial \tau^2} - \tau \sum_{j=1}^N a_{kj} \frac{\partial^3 v_{j\infty}(\tau)}{\partial \tau^3} \\ R &= 2 \sum_{j=1}^N a_{kj} \frac{\partial^2 v_{j\infty}(\tau)}{\partial \tau^2} - \frac{3!}{(1-v_{k\infty})^4} \left( \frac{\partial v_{k\infty}(\tau)}{\partial \tau} \right)^3 - \frac{4}{(1-v_{k\infty})^3} \frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} \frac{\partial v_{k\infty}(\tau)}{\partial \tau} \end{aligned}$$

Again rewritten as

$$\frac{1}{\tau(1-\nu_{k\infty})^2} \frac{\partial^3 \nu_{k\infty}(\tau)}{\partial \tau^3} - \sum_{j=1}^N a_{kj} \frac{\partial^3 \nu_{j\infty}(\tau)}{\partial \tau^3} = \frac{3}{\tau} \sum_{j=1}^N a_{kj} \frac{\partial^2 \nu_{j\infty}(\tau)}{\partial \tau^2} - \frac{3!}{\tau(1-\nu_{k\infty})^4} \left( \frac{\partial \nu_{k\infty}(\tau)}{\partial \tau} \right)^3 - \frac{6}{\tau(1-\nu_{k\infty})^3} \frac{\partial^2 \nu_{k\infty}(\tau)}{\partial \tau^2} \frac{\partial \nu_{k\infty}(\tau)}{\partial \tau}$$

leads to the matrix form

$$\mathcal{Q} \left( \frac{1}{\tau(1-\nu_{i\infty})^2} \right) \frac{\partial^3 V_{\infty}}{\partial \tau^3} = \frac{3}{\tau} A \frac{\partial^2 V_{\infty}}{\partial \tau^2} - \frac{6}{\tau} \text{diag} \left( \frac{1}{(1-\nu_{k\infty})^4} \right) \left( \frac{\partial V_{k\infty}(\tau)}{\partial \tau} \right)^3 - \frac{6}{\tau} \text{diag} \left( \frac{1}{(1-\nu_{k\infty})^3} \right) \frac{\partial^2 V_{k\infty}(\tau)}{\partial \tau^2} \frac{\partial V_{k\infty}(\tau)}{\partial \tau}$$

Introducing  $A \frac{\partial^2 V_{\infty}}{\partial \tau^2}$  from (A.7) as

$$A \frac{\partial^2 V_{\infty}}{\partial \tau^2} = \text{diag} \left( \frac{1}{\tau(1-\nu_{i\infty})^2} \right) \frac{\partial^2 V_{\infty}}{\partial \tau^2} - \frac{2}{\tau^2} \text{diag} \left( \frac{1}{(1-\nu_{i\infty})^2} \right) \frac{\partial V_{\infty}}{\partial \tau} + \frac{2}{\tau} \text{diag} \left( \frac{1}{(1-\nu_{k\infty})^3} \right) \left( \frac{\partial V_{k\infty}(\tau)}{\partial \tau} \right)^2 + \frac{2}{\tau^3} \text{diag} \left( \frac{1}{1-\nu_{i\infty}} \right) V_{\infty}$$

yields

$$\mathcal{Q} \left( \frac{1}{\tau(1-\nu_{i\infty})^2} \right) \frac{\partial^2 V_{\infty}}{\partial \tau^2} = R_3$$

where

$$R_3 = \frac{6}{\tau^4} \frac{V_{\infty}}{1-V_{\infty}} - \frac{6}{\tau^3} \frac{\frac{\partial V_{\infty}}{\partial \tau}}{(1-V_{\infty})^2} + \frac{3}{\tau^2} \frac{\frac{\partial^2 V_{\infty}}{\partial \tau^2}}{(1-V_{\infty})^2} + \frac{6}{\tau^2} \frac{\left( \frac{\partial V_{\infty}(\tau)}{\partial \tau} \right)^2}{(1-V_{\infty})^3} - \frac{6}{\tau} \frac{\left( \frac{\partial V_{\infty}(\tau)}{\partial \tau} \right)^3}{(1-V_{\infty})^4} - \frac{6}{\tau} \frac{\frac{\partial^2 V_{k\infty}(\tau)}{\partial \tau^2} \frac{\partial V_{k\infty}(\tau)}{\partial \tau}}{(1-V_{\infty})^3}$$

The computation illustrates the general structure (A.3) and demonstrates the Theorem.  $\square$

From a numerical point of view, the non-linear NIMFA steady-state matrix equation (A.4) only needs be solved once for a particular value of  $\tau$  so that the vector  $V_{\infty}(\tau)$  is known, as well as the generalized Laplacian  $\mathcal{Q} \left( \frac{1}{\tau(1-\nu_{i\infty})^2} \right)$ . The Taylor expansion

$$V_{\infty}(\tau + \Delta\tau) = \sum_{m=0}^{\infty} \frac{(\Delta\tau)^m}{m!} \frac{\partial^m V_{\infty}(\tau)}{\partial \tau^m} \quad (\text{A.8})$$

specifies the NIMFA infection probability vector  $V_{\infty}(\tau + \Delta\tau)$  at another effective infection rate  $\tau + \Delta\tau$ , provided that the Taylor series converges at  $\tau + \Delta\tau$ . As mentioned earlier in [122], unfortunately, the convergence radius of the series in (A.8) is difficult to determine in general. The left-hand side positive definite matrix  $\mathcal{Q} \left( \frac{1}{\tau(1-\nu_{i\infty})^2} \right)$  in (A.3) is the same for all orders  $m \geq 1$  and can be inverted if a high precision and many terms in the Taylor series (A.8) are required.

### A.5. THE VALUE OF $\tau_u$

We define  $\chi_D(\tau) = \sum_{i=1}^N \sum_{j=1}^{i-1} 1_{f_{ij}(V(\tau), D) < 0}$ , then the larger  $\chi_D$  is the higher the difference between the rankings of the infection probability at  $\tau$  and the nodal degree is. As shown in A.3, we plot  $\chi_D(\tau)$  as a function of the average fraction  $y_\infty$  of infected nodes for ER and BA random graphs with the average degree  $E[D] = 14$  as an example. We find that for both graphs  $\chi_D \approx 0$  when the average fraction  $y_\infty$  of infected nodes is above 0.9, which suggests that we can employ the value of  $\tau_u$  so that  $y_\infty(\tau_u) = 0.9$ . We have also done such tests on all the other networks in this paper and obtain the same conclusion. Hence, we employ the value  $\tau_u$ , leading to  $y_\infty(\tau_u) = 0.9$ , for all networks in this paper.

This choice of 0.9, though not necessarily optimal, is supported by the following aspects. Practically, we would like to choose  $y_\infty(\tau_u)$  as large as possible so that real-world prevalence levels are covered. Since real-world prevalence seldom reaches 0.9,  $y_\infty(\tau_u) = 0.9$  is large enough. Also, we would like to choose  $y_\infty(\tau_u)$  as large as possible so that  $\chi(\tau_c^{(1)} + \epsilon, \tau_u)$  well counts the total number of crossings. Moreover,  $y_\infty(\tau_u)$  should not be too large because the infection probability of the nodes are very close to each other when the prevalence is high, and the precision of numerical solution to compute the infection probability per node is not sufficient to distinguish nor to rank the nodes according to their infection probabilities. Furthermore, we observed that the crossing seldom happens when  $y_\infty(\tau) > 0.9$  in all the networks generated by the two network models as well as in real-world networks. This is due to the fact that the number of crossings decreases as  $\tau$  increases, as observed and discussed in the paper.

To compute the value of  $\tau_u$  which leads to a high prevalence (0.9), we can employ the Laurent series of the steady-state infection probability [122, 123]:

$$v_{i\infty}(\tau) = 1 + \sum_{m=1}^{\infty} \eta_m(i) \tau^{-m} \quad (\text{A.9})$$

where the coefficient  $\eta_1(i) = -\frac{1}{d_i}$  and

$$\eta_2(i) = \frac{1}{d_i^2} \left( 1 - \sum_{j=1}^N \frac{a_{ij}}{d_j} \right) \quad (\text{A.10})$$

and for  $m \geq 2$ , the coefficients obey the recursion

$$\eta_{m+1}(i) = -\frac{1}{d_i} \eta_m(i) \left( 1 - \sum_{j=1}^N \frac{a_{ij}}{d_j} \right) - \frac{1}{d_i} \sum_{k=2}^m \eta_{m+1-k}(i) \sum_{j=1}^N a_{ij} \eta_k(j)$$

Considering a large value of  $\tau_u$

$$v_{i\infty} = 1 - \frac{1}{\tau_u d_i} + O(\tau^{-2})$$

and

$$y_\infty = 1 - \frac{1}{\tau_u} E\left[\frac{1}{D}\right] + O(\tau^{-2})$$

then, ignoring the second order condition  $O(\tau^{-2})$ ,

$$\tau_u \approx \frac{1 - y_\infty}{E\left[\frac{1}{D}\right]}$$

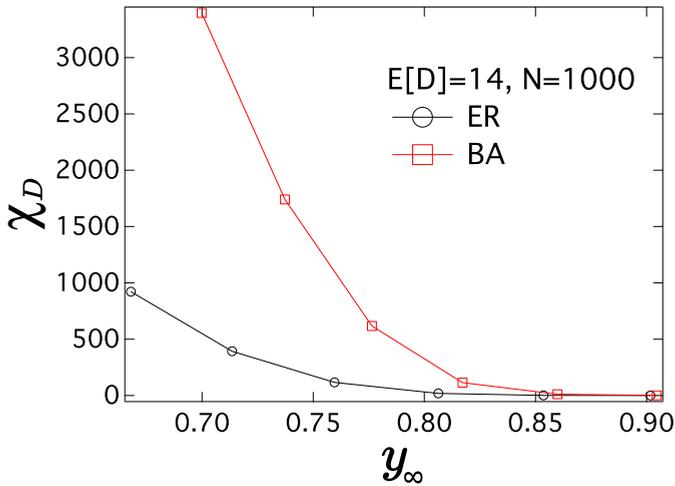


Figure A.3: The plot of  $\chi_D$  as a function of the average fraction of infected nodes. The results are averaged over 10 realizations

## A.6. REAL-WORLD GRAPHS

We use 6 connected and undirected graphs from real-world datasets. Some graphs are originally directed and may not be connected. We use the largest component of the unconnected graphs and change the directed graphs to undirected. The description of the 6 graphs are as follows:

1. GRQC: Arxiv GR-QC (General Relativity and Quantum Cosmology) collaboration network is from the e-print arXiv and covers scientific collaborations between authors papers submitted to General Relativity and Quantum Cosmology category. If an author  $i$  coauthored a paper with author  $j$ , there is link between  $i$  and  $j$ . The data covers papers in the period from January 1993 to April 2003.
2. NetSci: A coauthorship network of scientists working on network theory and experiment. The network was compiled from the bibliographies of two review articles on networks.
3. ODLIS: The network is based on the ODLIS: Online Dictionary of Library and Information Science (December 2000). The nodes are the terms in ODLIS and there is link between two terms if one is used to describe another one.
4. Roget<sup>2</sup>: The network contains cross-references in Roget's Thesaurus, 1879. Each node of the graph corresponds to one of the categories in the 1879 edition of Peter Mark Roget's Thesaurus of English Words and Phrases. There is a link between two categories if one is the reference of the other.

<sup>2</sup>See <http://vlado.fmf.uni-lj.si/pub/networks/data/dic/roget/Roget.htm>

5. Power: The network represents the topology of the Western States Power Grid of the United States.
6. Yeast: The protein-protein interaction network in budding yeast. There is link between protein  $i$  and protein  $j$  if they have the interaction.

In Table A.1, we list the size  $N$ , the average degree  $E[D]$ , the degree variance  $Var[D]$  and the normalized degree variance  $Var^*[D]$  of the 6 graphs. A larger size  $N$  always indicates a larger value of  $\chi_l$  in Fig. 8.2.

Table A.1: The real-world graph used in this paper.

	GRQC	NetSci	ODLIS	Roget	Power	Yeast
$N$	4158	379	2898	994	4941	2224
$E[D]$	6.46	4.82	11.30	7.32	2.67	5.94
$Var[D]$	74.42	15.46	679.61	23.66	3.21	63.70

## A.7. THE COMPARISON BETWEEN NIFMA AND THE CONTINUOUS-TIME SIMULATION

We compare the number of crossings obtained by NIMFA and the simulations of the exact SIS model. We show two examples of the comparison in Fig. A.4. Because the NIMFA epidemic threshold  $\tau_c^{(1)}$  is actually the lower bound of the real epidemic threshold, i.e.  $\tau_c^{(1)} < \tau_c$ , and to determine the value of  $\kappa_c = \frac{\tau_c}{\tau_c^{(1)}} > 1$  for different topology is difficult, we start the comparison from  $\kappa = 2$  (attempting to exclude the crossings near the epidemic threshold). Fig. A.4 shows that the results of the simulation and NIMFA agree with each other quite well for both networks when  $\kappa$  is not large. When  $\kappa$  is large, i.e. the infection probability of each node is high and close to each other, there might be some crossings caused by the limited precision of the numerical NIMFA solution or the simulations. Because the precision of the numerical solution is higher than that of the simulation of the exact SIS model, the number  $\chi$  of crossings obtained from the simulation tends to be larger than that from NIMFA if the effective infection rate  $\kappa$  is large.

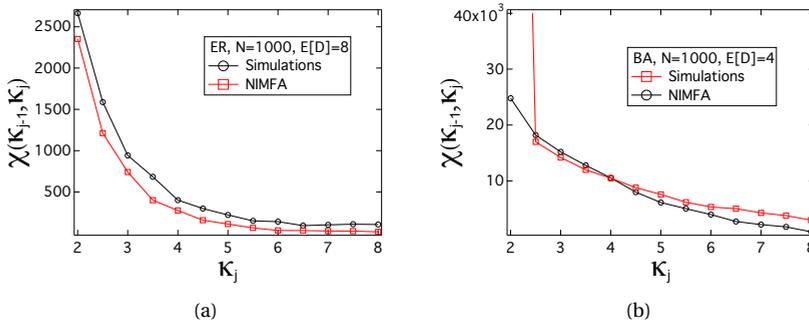
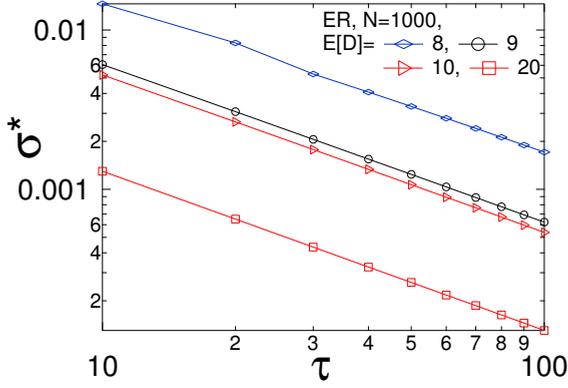


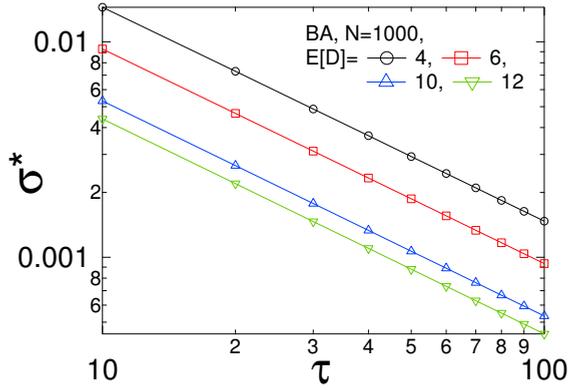
Figure A.4: The comparison of the number  $\chi$  of crossings between NIMFA and the simulation of the exact SIS model for (a) an ER random graph with  $N = 1000$  and  $E[D] = 8$ ; (b) a BA random graph with  $N = 1000$  and  $E[D] = 4$ . The linear sampling is employed with the step  $\Delta\kappa = 0.5$ .

### A.8. $\sigma^*$ AS A FUNCTION OF $\tau$

Here we show the relationship between  $\sigma^*$  and  $\tau$  when  $\tau \gg 1$ .



(a)

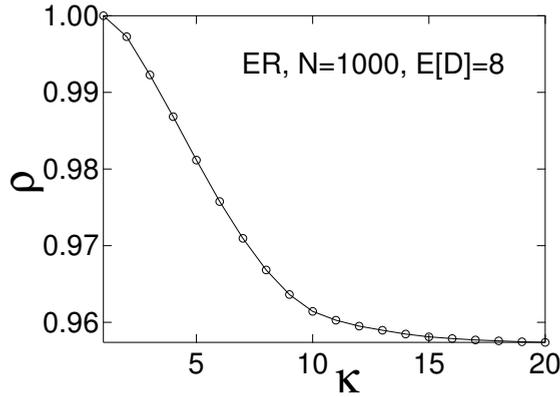


(b)

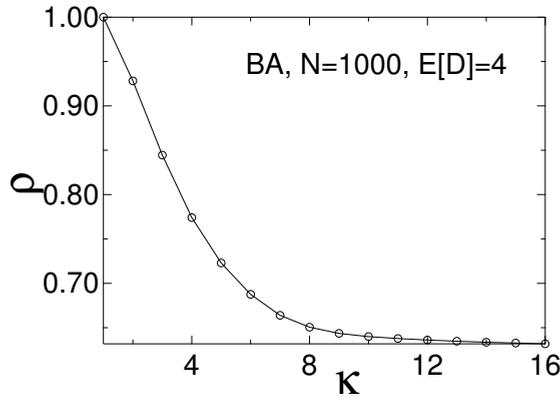
Figure A.5:  $\sigma^*$  as a function of  $\tau$  for (a) ER random graphs with the size  $N = 1000$  and (b) BA random graphs with the size  $N = 1000$ .

### A.9. THE SPEARMAN RANK CORRELATION $\rho$ AS A FUNCTION OF $\kappa$

In Fig. A.6, we choose two graphs, an ER random graph with  $E[D] = 8$  and a BA random graph with  $E[D] = 4$ , as the example. We plot Spearman Rank Correlation between  $V_\infty(\tau_c^{(1)} + \epsilon)$  and  $V_\infty(\kappa\tau_c^{(1)})$ . We find that the rank correlation decreases fast when the effective infection rate is small. Moreover, there tend to be a few nodes drastically changing ranks in BA random graphs but not in ER graphs.



(a)



(b)

Figure A.6: The Spearman rank correlation between  $V_\infty(\tau_c^{(1)} + \epsilon)$  and  $V_\infty(\kappa\tau_c^{(1)})$  as a function of  $\kappa$  for (a) ER random graphs with the size  $N = 1000$  and the average degree  $E[D] = 8$  and (b) BA random graphs with the size  $N = 1000$  and the average degree  $E[D] = 4$ .

# CURRICULUM VITÆ

## Bo Qu

05-03-1987      Born in Harbin, China.

### EDUCATION

1999–2005      Secondary School  
Xing Hua Junior High School, Harbin, China (1999–2002)  
Harbin Normal University High School, Harbin, China (2002–2005)

2005–2009      B.Sc. Information Security  
Training Reform Class, Shanghai Jiao Tong University (2005-2007)  
Information Security, Shanghai Jiao Tong University (2007-2009)

2009–2012      M.Sc. Computer Organization & Systems  
Lab of Cryptology and Computer Security, Shanghai Jiao Tong University

2012-2016      PhD.  
Delft University of Technology  
*Thesis:*          Dynamic Processes on Complex networks  
                         – The Role of Heterogeneity  
*Promotor:*      Prof. dr. A. Hanjalic  
*Copromotor:*   Dr. ir. H. Wang



# LIST OF PUBLICATIONS

11. **L. Liu, B. Qu, B. Chen, A. Hanjalic, and H. Wang**, *Modeling of Information Cascades in Social Media: An Empirical Research of WeChat Moments*, arXiv preprint arXiv:1704.03261.
10. **L. Liu, B. Chen, B. Qu, L. He, and X. Qiu**, *Data Driven Modeling of Continuous Time Information Diffusion in Social Networks*, IEEE Second International Conference on Data Science in Cyberspace, 2017.
9. **H. Wang, C. Chen, B. Qu, D. Li and S. Halvin**, *Epidemic Mitigation via Awareness Propagation in Communications Network: the Role of Time Scale*, Accepted by New Journal of Physics, to appear.
8. **B. Qu, C. Li, P. Van Mieghem and H. Wang**, *The nodal infection probability in SIS epidemic spreading*, Accepted by Scientific Report, to appear.
7. **B. Qu and H. Wang**, *SIS Epidemic Spreading with Heterogeneous Infection Rates*, IEEE Transactions on Network Science and Engineering, Issue 99, 2017.
6. **B. Qu and H. Wang**, *The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading*, International Workshop on Complex Networks and their Applications, 499-510, 2016.
5. **B. Qu and H. Wang**, *SIS Epidemic Spreading with Correlated Heterogeneous Infection Rates*, Physica A: Statistical Mechanics and its Applications 472, 13-24, 2016.
4. **B. Qu, A. Hanjalic, and H. Wang**, *Heterogeneous Recovery Rates against SIS Epidemics in Directed Networks*, NetGCoop 2014: International Conference on NETwork Games, COntrol and OPTimization. Trento, Italy.
3. **B. Qu, Q. Li, S. Havlin, H.E. Stanley and H. Wang**, *Nonconsensus opinion model on directed networks*, Physical Review E, 2014, **90**(5): 052811.
2. **Y. Liu, D. Gu, B. Li and B. Qu**, *Legitimate-reader-only attack on MIFARE Classic*, Mathematical and Computer Modelling, 2013, **58**(1): 219-226.
1. **B. Qu, D. Gu, Z. Guo, and J. Liu**, *Differential power analysis of stream ciphers with LFSRs*, Computers & Mathematics with Applications, 2013, **65**(9): 1291-1299.