Delft University of Technology

Performance of complex networks
Efficiency and Robustness

He, Zhidong

DOI
10.4233/uuid:a9ae2b62-0cf8-4cf3-86af-abd33bc99080

Publication date
2020

Document Version
Final published version

Citation (APA)
He, Z. (2020). *Performance of complex networks: Efficiency and Robustness* (1 ed.). [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a9ae2b62-0cf8-4cf3-86af-abd33bc99080

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.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

PERFORMANCE OF COMPLEX NETWORKS

EFFICIENCY AND ROBUSTNESS

PERFORMANCE OF COMPLEX NETWORKS

EFFICIENCY AND ROBUSTNESS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
17 March 2020 at 10:00 o'clock

by

Zhidong HE

Master of Science in Control Science
Zhejiang University, Hangzhou, China
born in Shanghai, China.

This dissertation has been approved by the promotor:

Prof. dr. ir. P. F. A. Van Mieghem

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. P. F. A. Van Mieghem	Delft University of Technology, promotor
Independent members:	
Prof. dr. ir. Lóránt A. Tavasszy	Delft University of Technology
Prof. dr. D. Li	Beihang University, China
Prof. dr. J. L. Marzo	University of Girona, Spain
Prof. dr. ir. R. E. Kooij	Delft University of Technology
Dr. J. L. A. Dubbeldam	Delft University of Technology
Dr. O. Cats	Delft University of Technology
Prof. dr. S. Hamdioui	Delft University of Technology, reserved

Keywords: Complex Networks, Network Robustness, Epidemic Spreading, Transport, Network Optimization

Printed by: ProefschriftMaken

Cover design by: Zhidong He

Published by: Zhidong He

Email: sgyyhzd@hotmail.com

Copyright © 2020 by Z. He

ISBN 978-94-6384-121-4

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

To my family

CONTENTS

Summary **xi**

Samenvatting **xiii**

1 Introduction **1**
- 1.1 Performance of networks 2
  - 1.1.1 Efficiency of Spread 3
  - 1.1.2 Robustness of Transport 4
- 1.2 Research questions 4
- 1.3 Outline 5
  - 1.3.1 Part I: Efficiency of spread on networks 5
  - 1.3.2 Part II: Robustness of transport in networks 6

Part I: Efficiency of spread on networks **7**

2 The spreading time in SIS epidemics on networks **9**
- 2.1 Introduction 10
- 2.2 Definition and determination of the spreading time 11
- 2.3 Distribution of the spreading time T_m 12
- 2.4 The average spreading time $E[T_m]$ in SIS processes 19
  - 2.4.1 Effect of the effective infection rate on $E[T_m]$ 19
  - 2.4.2 Effect of the shape parameter α on $E[T_m]$ 19
  - 2.4.3 Effect of the network size on $E[T_m]$ 21
- 2.5 Chapter summary 23

3 The fastest spreader in SIS epidemics on networks **25**
- 3.1 Introduction 26
- 3.2 The spreading time in epidemics on networks 27
- 3.3 The fastest spreader in SIS epidemics 27
  - 3.3.1 Change of the fastest spreader with τ in a barbell-like graph 28
  - 3.3.2 A heuristic topological metric for the fastest spreader 29
- 3.4 Numerical results 32
- 3.5 Chapter summary 35

4 Optimal induced spreading of SIS epidemics in networks **37**
- 4.1 Introduction 38
- 4.2 Preliminaries and Model 39
  - 4.2.1 SIS model in networks 39
  - 4.2.2 Heterogeneous NIMFA model 39

4.3 Static optimization for induced spreading 40
    4.3.1 Problem statements 40
    4.3.2 Global optimization by differential evolution 41
4.4 Dynamic optimization for induced spreading 42
    4.4.1 Problem statements 42
    4.4.2 The optimal solution. 42
4.5 Numerical results and discussion 43
    4.5.1 Numerical results in the static optimization 43
    4.5.2 Numerical results in the dynamic optimization 45
    4.5.3 Comparison between the static and the dynamic 49
4.6 Related work 50
4.7 Chapter summary. 52

5 Prevalence expansion in NIMFA **53**
5.1 Introduction 54
5.2 Expansion of the NIMFA steady-state prevalence 55
    5.2.1 NIMFA prevalence 55
    5.2.2 Expansion of the NIMFA steady-state prevalence 55
    5.2.3 Estimation of the radius of convergence 58
5.3 Radius of convergence R of the expansion in some specific graphs 58
    5.3.1 Complete graphs and regular graphs. 58
    5.3.2 Star graphs $K_{1,N-1}$ 58
    5.3.3 Path graphs P_N 60
    5.3.4 ER random graphs 60
    5.3.5 Scale-free graphs. 62
5.4 Effect of the topological properties on R 62
    5.4.1 Effect of the network topology on R 62
    5.4.2 Numerical tests 64
5.5 Chapter summary. 65

Part II: Robustness of transport on networks **67**

6 Topological approach to measure network recoverability **69**
6.1 Introduction 70
6.2 Topological approach for measuring network recoverability 71
    6.2.1 R-value and challenge 71
    6.2.2 Scenario A: recovery of any alternative link 71
    6.2.3 Scenario B: recovery of attacked links 73
    6.2.4 Comparison via envelopes and the recoverability indicators 73
6.3 Robustness metric and recovery strategy 74
    6.3.1 Robustness metrics 74
    6.3.2 Attack and recovery strategies 76
    6.3.3 Telecommunication networks 76

6.4 Results and discussion . . . 77
6.4.1 Envelope examples and comparison . . . 77
6.4.2 Comparison of recovery strategies . . . 77
6.4.3 Overview of the Link Ratio and the Energy Ratio . . . 79
6.4.4 Relation between Scenario A and Scenario B . . . 81
6.4.5 Impact of R-threshold . . . 81
6.5 Chapter summary . . . 83

7 Robustness assessment of multimodal freight transport network 85
7.1 Introduction . . . 86
7.2 The synchromodal network model . . . 87
7.2.1 Multi-layer network . . . 88
7.2.2 Origins and destinations (OD centroid) . . . 88
7.2.3 Interconnection . . . 89
7.2.4 Interdependency . . . 89
7.3 Network robustness assessment . . . 89
7.3.1 Link attributes . . . 89
7.3.2 Network perturbation . . . 90
7.3.3 Performance indicator . . . 91
7.3.4 Traffic assignment . . . 91
7.4 Case study: the Dutch container freight transport network . . . 93
7.4.1 Network model and configuration . . . 93
7.4.2 Robustness assessment under random failures . . . 96
7.4.3 Robustness assessment via node criticality . . . 96
7.4.4 Robustness assessment under capacity degradation . . . 98
7.4.5 Topological properties of critical nodes . . . 99
7.5 Chapter summary . . . 100

8 Optimization of convergence rate via algebraic connectivity 103
8.1 Introduction . . . 104
8.2 Algebraic connectivity for consensus processes in networks . . . 105
8.2.1 Consensus processes in undirected networks . . . 105
8.2.2 Consensus processes in directed networks . . . 106
8.3 Algebraic connectivity in undirected networks . . . 107
8.3.1 Topological perturbation for the algebraic connectivity . . . 107
8.3.2 Maximize the algebraic connectivity by adding links . . . 109
8.4 Generalized algebraic connectivity in directed networks . . . 112
8.4.1 Bounds of the generalized algebraic connectivity perturbation . . . 113
8.4.2 Maximize the generalized algebraic connectivity by adding links . . . 115
8.5 Numerical evaluations . . . 116
8.5.1 Undirected networks . . . 116
8.5.2 Directed networks . . . 118
8.6 Chapter summary . . . 119

9 Conclusion 121
9.1 Main contributions . . . 121
9.2 Directions for future work . . . 123

References 125
References 125

Acknowledgements 141

A Appendix for Chapter 2 143
A.1 Determination of the metastable state and the stability time 143
A.2 Simulation for a SIS process on networks 145
A.3 The generating function of the hitting time 146

B Appendix for Chapter 4 149
B.1 Proofs of Lemma and Theorem 149
B.1.1 Proof for Lemma 4.1 149
B.1.2 Proof for Theorem 4.1 149
B.2 Differential Evolution algorithm 150

C Appendix for Chapter 5 153
C.1 Recurrence of the coefficients $c_j(k)$ in the expansion of $v_{i\infty}(\tau)$ 153
C.2 Proof for Lemma 5.1 154
C.3 A lower bound of the radius of convergence R 154

Curriculum Vitæ 157

List of Publications 159

SUMMARY

Network performance is determined by the interplay of underlying structures and overlying dynamic processes on networks. This thesis mainly considers two types of collective dynamics on networks, spread and transport, which are ubiquitous in our daily lives, ranging from information propagation, disease spreading, to molecular motors on cytoskeleton and urban traffic. Exploring the approaches on optimizing the network performance is the fundamental motivation of this work, which helps to control processes on networks and to upgrade network-based services.

Although the properties of phase transition in Susceptible-Infected-Susceptible (SIS) processes have been investigated intensively, the time-dependent behavior of epidemics is still an open question. This thesis starts with the investigation of the *spreading time* (Chapter 2), which is the time when the number of infected nodes in the metastable state is first reached, starting from the outbreak of the epidemics. We observe that the spreading time resembles a lognormal-like distribution both for the Markovian and the non-Markovian infection processes.

As a follow-up work of Chapter 2, we identify the fastest initial spreaders with the shortest average spreading time in epidemics on a network, which helps to ensure an efficient spreading (Chapter 3). We show that the fastest spreader changes with the effective infection rate of a SIS epidemic process, which means that the time-dependent influence of a node is usually strongly coupled to the dynamic process and the underlying network. We propose the *spreading efficiency* as a metric to quantify the efficiency of a spreader and identify the fastest spreader, which is adaptive to different infection rates in general networks.

For maximizing the utility of spread, we introduce *induced spreading*, which aims to maximize the infection probabilities of some target nodes by adjusting the nodal infection rates (Chapter 4). We assume that the adjustment of the nodal infection rates has an associated cost and formulate the induced spreading for SIS epidemics in networks as an optimization problem under a constraint on the total cost. We address both a static model and a dynamic model for the optimization of the induced SIS spreading. We show that the infection rate increment on each node is coupled to both the degree and the average hops to the target nodes in the static optimization method. In the dynamic method, the effective resistance is a good metric to indicate the minimum total cost for targeting a single node.

The average fraction of infected nodes in the NIMFA steady state, also called the steady-state prevalence, in terms of the effective infection rate can be expanded into a power series around the NIMFA epidemic threshold. Practically, we can faster compute the nodal infection probability of the NIMFA steady-state by the truncated expansion with enough terms and an effective infection rate within the radius of convergence. Thus, we investigate the radius of convergence that validates the Taylor expansion of the steady-state prevalence in Chapter 5. We show that the radius of convergence of the steady-state prevalence expansion strongly depends upon the spectral gap of the adjacency matrix.

The research on the robustness of transport on networks mainly encompasses two robustness assessment approaches, along with their applications in communication networks and freight transport networks, respectively. *Network recoverability* refers to the ability of a network to return to a desired performance level after suffering malicious attacks or random failures (Chapter 6). We propose a general topological approach and recoverability indicators to measure the network recoverability in two scenarios: 1) recovery of damaged connections and 2) any disconnected pair of nodes can be connected to each other. By applying the effective graph resistance and the network efficiency as robustness metrics, we employ the proposed approach to assess 10 real-world communication networks. For vehicle transport systems, Chapter 7 proposes a robustness assessment for *multimodal transport networks*. The representation of interdependent networks is an excellent proxy for the structure of multimodal transportation systems. We apply our robustness assessment model to the Dutch freight transport, taking into account three modalities: waterway, road and railway. The node criticality, defined as the impact of a node removal on the total travel cost, resembles a power-law distribution, which implies scale-free property of the robustness against infrastructure disruptions.

Many transport processes have a similar objective that all nodes reach an agreement regarding a certain quantity of interest by exchanging the nodal states with their neighboring nodes, which are described by the *consensus model* in networks (Chapter 8). The robustness of consensus processes is related to the convergence speed to the stability under external perturbations. The (generalized) algebraic connectivity of a network characterizes the lower-bound of the exponential convergence rate of consensus processes. We investigate the problem of accelerating the convergence of consensus processes by adding links to the network. We propose a greedy strategy for undirected network and further extend our approach to directed networks. Numerical tests verify the better performance of our methods than other metric-based approaches.

This thesis considers two dynamic processes on networks and covers performance analysis and optimizations, by means of problem proposal, theoretical analysis, case study and algorithm designing. The developed concepts related to network efficiency and robustness provide a better understanding of collective dynamics on complex networks. The applicability of our methodologies bridges theoretical network models and realistic applications, as well as demonstrates the promising efficacy of network science.

SAMENVATTING

De prestatie van een netwerk wordt bepaald door de interactie tussen de onderliggende structuur van het netwerk en de dynamische processen die plaatsvinden op het netwerk. Dit proefschrift beschouwt twee typen dynamiek die plaatsvinden op een netwerk: verspreiding en transport. Beiden zijn alom vertegenwoordigd in het dagelijkse leven, van ziekteverspreiding en voortplanting van informatie tot het stadsverkeer en verplaatsing van motoreiwitten over cytoskeletten op moleculair niveau. Het verkennen van methoden die de prestatie van een netwerk optimaliseren is de belangrijkste motivatie voor dit werk, dat zijn methoden die netwerkprocessen aansturen en netwerkservices kunnen verbeteren.

De eigenschappen van de faseovergang in het vatbaar-geïnfecteerd-vatbaarproces (Engels: susceptible-infected-susceptible process, afgekort het SIS-proces) is uitvoerig bestudeerd in de literatuur, maar het doorgronden van het tijdsveranderlijke gedrag van epidemieën is grotendeels nog een open vraagstuk. In dit proefschrift beginnen we met het onderzoeken van de epidemische verspreidingstijd (Hoofdstuk 2), de tijd van de uitbraak van een epidemie tot het totaal aantal besmette individuen voor het eerst gelijk is aan het gemiddelde aantal besmette individuen in de metastabiele toestand. Wij merken op dat de verdeling van de verspreidingstijd lijkt op een lognormale verdeling voor zowel Markoviaanse en niet-Markoviaanse verspreidingsprocessen.

Vervolgens identificeren wij de snelste initiële verspreiders in een netwerk met behulp van de kortste gemiddelde verspreidingstijd (Hoofdstuk 3). Hiermee kan een efficiënte verspreiding gegarandeerd worden als dit de doelstelling is. Wij laten zien dat de snelste verspreider verandert als de effectieve infectiesnelheid van het SIS-proces verandert, wat een indicatie is dat de tijdsafhankelijke invloed van een individu sterk gerelateerd is aan het dynamische proces en het onderliggende netwerk. Tot slot introduceren wij de *spreidingsefficiëntie* als maat voor de efficiëntie van een verspreider en gebruiken deze om de snelste verspreider te kunnen identificeren.

Voor de optimalisatie van een verspreiding introduceren wij het concept *geïnduceerde verspreiding*, een techniek die de besmettingskansen van individuen maximaliseert door het aanpassen van individuele infectiesnelheden (Hoofdstuk 4). Hierbij modelleren wij dit proces met kosten voor het aanpassen van deze individuele infectiesnelheden en formuleren wij de geïnduceerde verspreiding voor SIS-epidemieën als een optimalisatieprobleem met een restrictie op de totale kosten. We beschouwen zowel een statisch als een dynamisch model voor de optimalisatie van de geïnduceerde verspreiding. We laten zien dat de toename in individuele infectiesnelheid is gerelateerd aan zowel de graad van de knoop van het individu in het netwerk als de gemiddelde afstand tot de knopen van de doelwitten van de geïnduceerde verspreiding. Voor de dynamische methode blijkt de effectieve graafweerstand een goede maatstaf om te identificeren bij welk individu geïnduceerde besmetting de totale kosten minimaliseert.

De stabiele toestandsprevalentie, het gemiddelde percentage geïnfecteerde knopen in de NIMFA stabiele toestand, kan worden uitgedrukt als een machtreeks rond de epidemische kritieke drempel. Dit heeft als voordeel dat infectiekansen in de NIMFA stabiele toestand sneller uitgerekend kunnen worden zolang er genoeg termen gebruikt worden in de benadering en zolang de effectieve infectiesnelheid binnen de convergentiestraal van de reeks ligt. Wij onderzoeken de convergentiestraal die de Taylorreeks van de stabiele toestandsprevalentie van Hoofdstuk 5 valideert. Wij laten zien dat de convergentiestraal sterk afhangt van de grootte van het verschil tussen de eerste twee eigenwaarden van de verbindingsmatrix van het netwerk.

Het onderzoek naar de robuustheid van transport over netwerken richt zich met name op twee methoden voor het meten van robuustheid en op hun respectievelijke toepassingen in communicatienetwerken en vervoersnetwerken. *Netwerkherstelbaarheid* verwijst naar de mate waarin een netwerk kan terugkeren naar het gewenste prestatieniveau nadat het bloot heeft gestaan aan vijandelijke aanvallen of heeft geleden onder storingen (Hoofdstuk 6). Wij introduceren een algemeen topologisch raamwerk met verschillende indicatoren om de herstelbaarheid van een netwerk te meten in twee verschillende herstelscenario's: bij het herstellen van een beschadigde verbinding herstellen we in scenario 1) de verbinding die is beschadigd en in scenario 2) een verbinding tussen twee willekeurig gekozen knopen in het netwerk. Door gebruik te maken van de effectieve graafweerstand en de netwerkefficiëntie als robuustheidsindicatoren analyseren wij de robuustheid van 10 bestaande communicatienetwerken. In Hoofdstuk 7 beschouwen we de robuustheid van voertuigtransport in *multimodale vervoersnetwerken*. Onderling-afhankelijke netwerken zijn een uitstekende benadering van de structuur van multimodale vervoerssystemen. Wij passen ons model voor robuustheid toe op het Nederlandse vrachttransportnetwerk met drie vervoerswegen: water, weg en spoor. Het belang van een knoop is gedefinieerd als de impact op de totale vervoerskosten als gevolg van het verwijderen van die knoop. Wij tonen aan dat de verdeling van het belang van de knopen een machtsverband (Engels: power law) vertoont, wat impliceert dat de robuustheid met betrekking tot infrastructurele verstoringen schaalvrij (Engels: scale-free) is.

Veel transportnetwerken hebben de doelstelling om alle knopen overeenstemming te laten bereiken in termen van een bepaalde grootheid door het uitwisselen informatie met aangelegen knopen, een proces wat wordt beschreven door het *consensusmodel* (Hoofdstuk 8). De robuustheid van consensusprocessen is gerelateerd aan de convergentiesnelheid en stabiliteit onder externe perturbaties. De algebraïsche connectiviteit van een netwerk vormt een ondergrens voor de exponentiele convergentiesnelheid van consensusprocessen. Wij onderzoeken hoe de convergentie versneld kan worden door het toevoegen van verbindingen in het netwerk. Wij introduceren een greedy strategie voor ongerichte netwerken en we breiden onze aanpak uit naar gerichte netwerken. Numerieke tests wijzen uit dat onze methode beter presteert dan vergelijkbare methoden die ook op netwerkgerelateerde maten gebaseerd zijn.

Dit proefschrift onderzoekt twee dynamische processen op netwerken en behandelt prestatieanalyse en optimalisaties in deze processen door middel van probleemidentificatie, theoretische analyse, casussen en ontwerp van algoritmen. De ontwikkelde concepten gerelateerd aan netwerkefficiëntie en robuustheid verschaffen een beter inzicht in de dynamiek in processen op netwerken. De toepasbaarheid van onze methoden verbindt theoretische netwerkmodellen met het dagelijkse leven en toont bovendien de toegevoegde waarde van de netwerkwetenschap aan.

1

INTRODUCTION

“Nature flies from the infinite, for the infinite is unending or imperfect, and Nature ever seeks an end.”

— Generation of Animals

NETWORKS are ubiquitous in the world and in our daily lives. Physically, networks can represent vastly different objects including taffic infrastructures with roads and junctions [1], power grids with wires and substations [2], the Internet with fiber and switches [3], neural systems with axons and neurons [4]. Abstractly, networks can describe friendship or collaboration relations among individuals [5], social contacts and following connections in social media [6], the way the functional brain works [7] and so on. Network is a prototype consisting of a collection of agents with connections where various phenomena emerge due to the complexity of dynamics on networks. Network model becomes a tool for interdisciplinary research, which includes statistical mechanism, particle physics, computer science, electrical engineering, biology, economics, ecology and sociology.

It is universally recognized that network science origins from graph theory. In 1736, Leonhard Euler (1707-1783) solved the Seven Bridges of Königsberg problem, which is regarded as the originating point of graph theory. The investigation on the *network structure* successively unveils topological properties and their practical implications. Erdős–Rényi model [8] established the random graph theory in 1959. Besides introducing probability theory into graph theory, random graph model opened doors of physics on general networks, e.g., percolation theory, which previously based on lattices in Euclidean spaces. In 1998, Watts and Strogatz [9] proposed a model to generate small-world networks, which complements theoretical basis for the phenomenon known as "six degree of separation". In 1999, Albert and Barabási [10] discovered scale-free property in the Internet and the World Wide Web, where the degree distribution follows a power law. The Barabási–Albert model proposed the method that generates scale-free networks which introduces the preferential attachment mechanism in network growth.

Beyond network typologies, *network dynamics* is generally recognized as the macroscopic emergent phenomena due to the collective dynamics of microscopic individual behaviors. We can understand network dynamics in another positivism view. We quantify the practical significance of a kinetic model by the probability that the reality agrees with the behavior of this model

$$\Pr[\text{Reality} \xrightarrow{\text{agrees with}} \text{Model}]$$

Traditional deterministic models generally feature the fundamental behaviors of dynamics, but ignore the practical limitations of communication between individuals, e.g. geographic restriction, interpersonal relationship and communication medium. Physically, the network topology essentially limits the range of interactions among individuals. Considering that the network is the constraint or the condition rather than the background, network science aims to explore the applicability of dynamic models in network cases, which means to improve the conditional probability

$$\Pr[\text{Reality} \xrightarrow{\text{agrees with}} \text{Model}|\text{Networks}]$$

In these 280 years, such network constraints are introduced to a large bundle of traditional dynamics, e.g., synchronization [11], spread [12], percolation [13], which are previously rooted in end-to-end, well-mixed [14] and lattice cases. The dramatic impacts of the underlying network on dynamic processes are addressed, which helps to describe the realistic collective dynamics better and to inspire many related areas of science.

1.1. PERFORMANCE OF NETWORKS

Network science has entered a new stage, where more and more researchers are devoting themselves to practical applications of network approaches and aim to improve the performance of networked systems. The performance of a network is defined by the interplay of the structure of the network and the dynamic process that runs on top of the network (illustrated by Figure 1.1). The structure of the network, represented by the underlying topology, defines the relations (represented by links) among individuals (represented by nodes). The dynamics on the network usually refer to collective behaviors of individuals in a spontaneous way or driven by external operations. The investigation on the performance of networks is significant for not only predicting the trend of the dynamic process but also providing better network-based services.

A large number of metrics of network performances can be classified into two aspects: *efficiency* and *robustness*. The efficiency of networks is the ability to avoid wasting materials, energy, money, and time in producing a desired output or providing a desired service. Efficiency of networks, putting forward higher demands than effectiveness, usually concerns various optimization problems to minimize the budgets or maximize the utility, e.g., pinning complex networks via controlling a minimum number of nodes [15], maximizing the influence of spread with a fix number of initial spreaders [16]. The optimization problems on complex networks are usually extremely high dimensional,

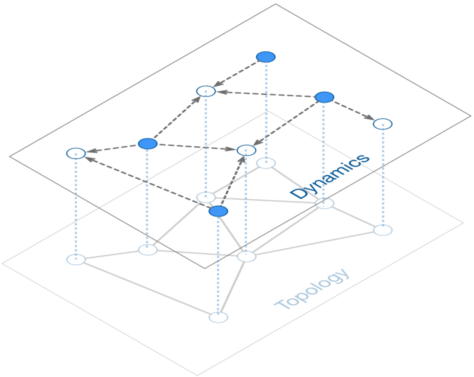


Figure 1.1: Illustration of framework for analyzing network performances.

which obstacles us to apply traditional approaches and motivates us to propose novel methods by making use of properties of complex networks. On the other performance metric, the robustness is interpreted as the maintenance of functionality under external perturbations [17]. Robustness is a long-lived research topic in engineering fields, but the robustness assessment of complex networks still eludes us in recent years. In a network perspective, Van Mieghem *et al.* [17] propose a framework, the R-model, for quantifying topological network robustness by considering both a network topology and a service for which the network is designed. Albert *et al.* [18] emphasized the crucial influence of the underlying topology on the complex communication networks, i.e., scale-free networks present high tolerance to random failures but are vulnerable to targeted attacks. Broadbent *et al.* [19] proposed the percolation model to analytically study the robustness of networks, which was followed by several further studies on cascading failures [20][21]. The analytical framework proposed by Gao *et al.* [22] collapsed the behaviour of different networks onto an universal resilience function by separating the roles of the system's dynamics and topology.

Two sorts of ubiquitous dynamic processes, i.e. spread and transport, are taken into account in this thesis.

1.1.1. EFFICIENCY OF SPREAD

The first concerned process in this thesis is *epidemics*, also called non-conserved *spread* [23], whose model was first proposed by Jacob Bernoulli [24]. Spreading processes can describe the transmission of most infectious diseases, neural excitation, information and rumors propagation [12]. The individuals in a spread are divided into several compartments, e.g., the susceptible, the infectious, or the recovered. The individual states switch among these compartments both by contacts between individuals and by spontaneous processes. Previous research has provided an exhaustive investigation on the phase transition behaviors and estimation of prevalence for different epidemic models [12] [25]. The N-intertwined mean-field approximation (NIMFA) proposed by Van Mieghem [26] is a reasonably accurate approximation of the exact epidemics on a network [27].

The further study on optimizing spreading processes encounters two variants corresponding to the underlying topology and the dynamics. The first variant is the adjustment on the underlying topology of the network to control/eliminate a virus spreading [28] by making use of the fact that the epidemic threshold is strongly related to the largest eigenvalue of the underlying network. The other variant is the adjustment on the spreading process, which refers to the optimization of the heterogeneous individual infection behaviors (e.g. the infection rates and the curing rates) to maximize/minimize the influence of a spread [29]. The second variant includes another research problem on the identification of the influential initial spreaders [30]. Rather than the static performance of spreading processes in the steady-state which has been investigated in most previous works, this thesis focuses more on the time-dependent performance. The issues such as investigation on the spreading time, identifying the fastest initial spreader, dynamic allocation of spreading resources will be discussed in this thesis. All our efforts aim to promote a more *efficient spread* on the network.

1.1.2. ROBUSTNESS OF TRANSPORT

The other process concerned in this thesis is *transport*. Different from the epidemic models where the virus can reproduce in the system, most transport phenomena are grounded in two primary concepts: the conservation laws and the constitutive equations [31]. Transport processes cover a wide range of real-world dynamics, e.g., fluid flow in tanks [32], power transmission in smart grid, packets delivery by optical fiber [33], vehicles driving in transportation networks [1].

Transport networks are prone to suffer from various perturbations, such as infrastructure failures and malicious attacks, which highlights the significance of the robustness performance of transport on networks. Unfortunately, the definition of robustness and the approach for robustness assessment still outstanding issues without consensus. Taking performance of specific transportation services into account, the robustness of data communication networks should reflect packets end-to-end readability; the robustness of road networks emphasizes the demands of the travel time and the congestion/free state; the robustness of networked control systems against external perturbations refers to the convergence rate to a stability [34]. The diversity in various robustness requirements challenges the development of a generalized robustness improvement framework. Further, the research gaps in optimization for network robustness encompasses two aspects. Firstly, improved robustness assessment models need to be proposed to evaluate the robustness performance of real-world structures, while new structural types and application scenarios are emerging and updating. Secondly, the optimal strategy for improving the network robustness could be theoretically intractable in some cases (e.g., in directed networks), which requires original feasible and heuristic methods to upgrade the network robustness effectively.

1.2. RESEARCH QUESTIONS

This thesis is motivated by the goal to promote promising applications of network theory in the real world. This thesis encompasses analyzing behaviors of dynamics, proposing performance assessment approaches, addressing network optimization problems, and verifying our approaches by case study. Some of the research challenges considered in this thesis are the following:

Can characteristic time help us to investigate the time-dependent behavior of epidemics process in networks? If we define the spreading time as the time when a spreading process reaches the meta-stable state, how is the distribution of the spreading time in stochastic spreading processes? What is the implication of the distribution of spreading time? (Chapter 2)

How can we further identify the fastest initial spreader for an efficiency spreading process with the shortest average spreading time? We need to understand what factors can influence the topological property of this fastest spreader. (Chapter 3)

If we aim to guide the virus to some determined regions instead of the whole network, how can the resource be allocated on nodes if the total budget is limited? Does the dynamic optimization benefit the targeted (induced) spreading on networks? What heuristic algorithm is feasible for high-dimensional network optimization problems? How is the cost scaling of resource for targeted spreading? (Chapter 4)

Since the prevalence in the steady state in the mean-field model can be represented by a Taylor series, we are curious about how to determine the validate range (radius convergence) of this series. (Chapter 5)

Relative to robustness against the failures, how can we assess the recoverability of a network? How is the recoverability performance of real communication networks? (Chapter 6)

How can we describe the multimodal transport system as a network model? How can we evaluate the robustness of multimodal transport networks under attacks and failures? How is the robustness performance of the Dutch freight transport network? (Chapter 7)

What topological property determines the convergence rate of consensus processes in undirected or directed networks? What is the effective strategy of adjusting the underlying topology to improve the convergence rate? (Chapter 8)

This thesis dedicates to a better understanding and solutions of the above mentioned questions.

1.3. OUTLINE

This thesis is organized into two parts as Figure 1.2. Part I focuses on the epidemic processes and the method to improve the efficiency performance of spread on networks. Part II investigates the robustness of transport network.

1.3.1. PART I: EFFICIENCY OF SPREAD ON NETWORKS

Chapter 2 investigates the spreading time in a Susceptible-Infected-Susceptible (SIS) process, which is the time when the number of infected nodes in the metastable state is first reached, starting from the outbreak of the epidemics.

Chapter 3 proposes the method to identify the fastest spreaders, with the shortest average spreading time, in epidemics on networks, which helps to ensure an efficient spreading. We show how the fastest spreader changes with the effective infection rate.

Chapter 4 introduces a new application of epidemics, induced spreading, which aims to maximize the infection probabilities of some target nodes by adjusting the

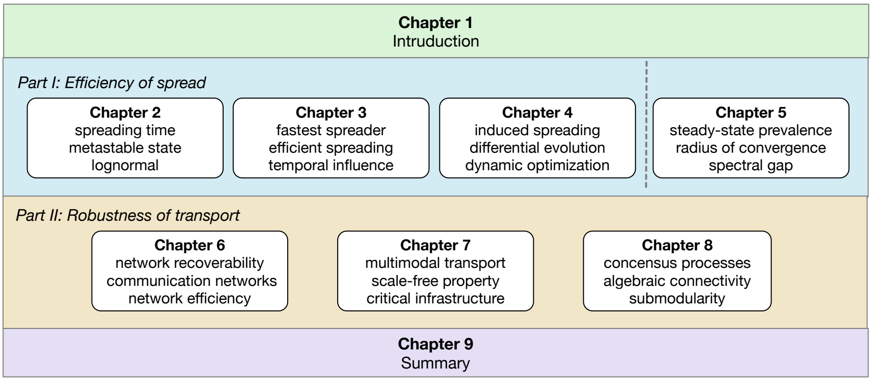

Chapter 1
Intruduction
Part I: Efficiency of spread
Chapter 2
spreading time
metastable state
lognormal
Chapter 3
fastest spreader
efficient spreading
temporal influence
Chapter 4
induced spreading
differential evolution
dynamic optimization
Chapter 5
steady-state prevalence
radius of convergence
spectral gap
Part II: Robustness of transport
Chapter 6
network recoverability
communication networks
network efficiency
Chapter 7
multimodal transport
scale-free property
critical infrastructure
Chapter 8
concensus processes
algebraic connectivity
submodularity
Chapter 9
Summary

Figure 1.2: Schematic depiction of the thesis.

nodal infection rates. The strategies for induced spreading are proposed for both static optimization and dynamic optimization.

Chapter 5 focuses on the average fraction of infected nodes in the NIMFA steady state, which can be expanded into a power series in terms of the effective infection rate around the NIMFA epidemic threshold. We determine the radius of convergence of this Taylor series, and investigate the relation between underlying topology and the radius of convergence.

1.3.2. PART II: ROBUSTNESS OF TRANSPORT IN NETWORKS

Chapter 6 proposes a general topological approach and recoverability indicators to measure the network recoverability in two scenarios. By applying the effective graph resistance and the network efficiency as robustness metrics, we employ the proposed approach to assess 10 real-world communication networks.

Chapter 7 proposes an approach on network modeling and robustness assessment for multimodal freight transport networks, which captures the features of interconnection and interdependency. We apply our robustness assessment model to the Dutch freight transport and identify the most critical infrastructure.

Chapter 8 investigates the problem of accelerating the convergence of consensus processes by adding links to the network. We propose a greedy strategy for adding links in undirected networks, and extend our strategy to directed networks.

Chapter 9 summarizes the contributions of this thesis and discusses some future work.

PART I: EFFICIENCY OF SPREAD ON NETWORKS

2

THE SPREADING TIME IN SIS EPIDEMICS ON NETWORKS

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

This chapter is based on the published paper [35].

2.1. INTRODUCTION

EPIDEMIC spreading on networks is a ubiquitous process, which can describe the information spreading on social networks [12], emotions [36], biological diseases [26] and failures in networked systems [37]. The Susceptible-Infected-Susceptible (SIS) model is a simple epidemic model where each infected item can be cured, and becomes susceptible again after recovering from the disease. Since the epidemic is a time-dependent spreading process, we are naturally concerned with characteristic times that can be applied to predict or control the spreading process. In spite of the simplicity of the SIS process, unfortunately, only a few results for exact SIS times on a generic graph have been presented [38, p. 460].

In the Susceptible-Infected-Susceptible (SIS) epidemics on a graph, the ratio between the infection rate β and the curing rate δ is called the effective infection rate $\tau = \beta/\delta$. The SIS model features a phase transition [39] around the epidemic threshold τ_c. Viruses with an effective infection rate τ above the epidemic threshold τ_c can infect a sizeable portion of the population on average and stay for a long time in the network. This long period is called the metastable state. Specially, in the Markovian SIS model, the infection processes and the curing processes are Poissonian. A first-order mean-field approximation of the epidemic threshold $\tau_c^{(1)} = 1/\lambda_1(A)$, where $\lambda_1(A)$ is the spectral radius of the adjacency matrix A, was shown [26][40] to be a lower bound for the epidemic threshold, $\tau_c^{(1)} < \tau_c$.

Due to the existence of an absorbing state, which is the overall healthy or disease-free state in the SIS process, any initial infection will ultimately extinguish in any finite graph. The time until the network reaches the all-healthy state is called the extinction time, or alternatively, the time to absorption or the survival time [41]. When the effective infection rate τ is below the epidemic threshold τ_c, the infectious process dies out exponentially fast [42][43], which is called quick die out or *early* extinction. A sufficient condition for *slow* die out [44] is that the effective infection rate τ is above the epidemic threshold τ_c. If the effective infection rate $\tau > \tau_c$, the infection stays very long on average in any sufficiently large network [45]. The average survival time is dominated by the second largest eigenvalue of the infinitesimal generator of the Markov chain [41][46].

In real-world large graphs, the extinction time is much longer than the actually observed time that an epidemic lasts. Therefore, besides the extinction time, we are interested in characteristic times before the absorbing state is reached. Van de Bovenkamp and Van Mieghem [47] showed that the average hitting time to the metastable state can be computed by using a uniformed embedded Markov chain for the complete graph and the star graph. The modified SIS model in [47] removes the absorbing state directly, implying that the process prevents itself from extinction and restarts to reach the metastable state, from one infected node. Thus, the average time to the metastable state is slightly overestimated, because the restarted process with one infected node usually needs a longer time to reach the metastable state.

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

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

This chapter is organized as follows. Section 2.2 introduces the definition and determination of the spreading time. We investigate the distribution of the spreading time T_m in Section 2.3. In Section 2.4, we further present the effect of the effective infection rate τ, the non-Markovian infection times and the network size N on the average spreading time. We conclude the chapter in Section 2.5. We define the metastable state and the stability t_s in a SIS process in Appendix A.1. Appendix A.2 presents the procedure of the simulator for SIS epidemics (SSIS). The generating function of the hitting time is derived in Appendix A.3.

2.2. DEFINITION AND DETERMINATION OF THE SPREADING TIME

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

Definition 2.1 *In an epidemic process, the metastable state is reached at the stability time* t_s*, which is the smallest time obeying* $\frac{d\overline{y}(t)}{dt}\big|_{t>t_s} < \epsilon$*, where the average fraction of infected nodes is* $\overline{y}(t) = \frac{1}{N}E[I(t)]$*, with* $I(t) \geq 1$ *is the number of infected nodes at time* t*, and* ϵ *is a small positive real number that needs to be agreed upon.*

A more detailed discussion on the determination of the stability time is presented in Appendix A.1.

Definition 2.2 *The spreading time* T_m *is defined as the first time when the number* $I_m = I(t_s)$ *of the infected nodes in the metastable state is reached, starting from one initially infected node.*

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

$$\Pr[T_m \leq t] = \sum_{n=1}^{N} \Pr[T_m \leq t | I(t_s) = n] \Pr[I(t_s) = n] \tag{2.1}$$

Thus, the average spreading time $E[T_m]$ follows from (2.1) as

$$E[T_m] = \sum_{n=1}^{N} E[T_{H_n}] \Pr[I(t_s) = n] \tag{2.2}$$

where the hitting time $T_{H_n} = T_m\big|_{I(t_s)=n}$ is the first time when the process reaches the state with n infected nodes. After differentiating both sides of (2.1) with respect to t, we obtain the probability density function (pdf) of the spreading time $f_{T_m}(t)$:

$$f_{T_m}(t) = \sum_{n=1}^{N} f_{T_m}(t|I(t_s) = n)\Pr[I(t_s) = n] \tag{2.3}$$

Physically, the spreading time T_m describes the spreading velocity in the early stage of the spreading process, which depends on the local topology around the initial spreaders. After T_m time units, the epidemic approximates the metastable state and already infected a substantial part of the population. Thus, the action of control is preferred to be taken earlier than the average spreading time $E[T_m]$. The average spreading time together with the expected number of infected individuals in the metastable state can guide public health officials in establishing the amount of resources and the available time for the implementation of their mitigation strategies.

Due to the limitation of the analytical methods, an event-driven simulator SSIS (see Appendix A.2) for the SIS spreading process on a network is implemented based on the Gillespie algorithm [48] to estimate the spreading time. For an unaltered graph and a fixed effective infection rate τ, the epidemic begins with one initially infected node and lasts for the period of t_{limit} time units which is ensured to be long enough to make the spreading process reach the metastable state but not the absorbing state. We record every time point t_k when the kth event happens, as well as the corresponding number of the infected nodes $i(t_k)$ immediately after the kth event. Assume that $0 < t_1 \le t_2 \le \cdots \le t_m < t_{limit}$, then m events have occurred on the timeline before the time limit t_{limit}. After identifying the metastable state and the stability time t_s (see Appendix A.2), we then determine the spreading time t_m in each realization. The spreading time can be determined from the time t_m when the number of infected nodes $i(t_m)$ first equals to the number $i(t_s)$ of infected nodes at the stability t_s of the metastable state. The random variable T_m corresponds to the spreading time t_m in all realizations that do not go extinct. Figure 2.1 illustrates the estimation scheme of the spreading time t_s in a complete graph K_{50}, which also shows the Gaussian-like distribution of the number of infected nodes in the metastable state.

2.3. DISTRIBUTION OF THE SPREADING TIME T_m

We first investigate the distribution of the spreading time T_m in the Markovian SIS process. The hitting time T_{H_i} is the first time when the Markov process reaches the state with i infected nodes, starting from one initial spreader. The epidemic process in the complete graph K_N is a birth and death process. Assume that the time is measured in units of $1/\delta$, the average hitting time $E[T_{H_i}]$ from one initial spreader can be analytically derived [47] as

$$E[T_{H_i}] = \sum_{j=1}^{i-1} \sum_{k=0}^{i-j-1} \frac{(N-i+k)!\tau^{j+k-i}}{j(N-j)!}. \tag{2.4}$$

in the modified SIS (MSIS) model [49], where the absorbing state is removed in MSIS Markovian chain. However, a hitting time analysis is tractable when the spreading process can be described as a simple, analytically tractable Markov chain [47].

Figure 2.2 exemplifies the average hitting time $E[T_{H_i}]$, from one initial spreader, as a function of the fraction $y = \frac{i}{N}$ of the infected nodes in the complete graph K_{50}

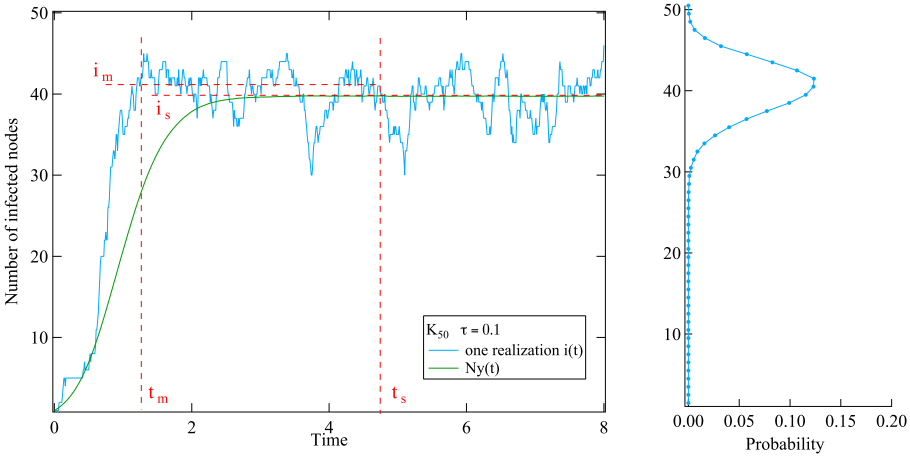


Figure 2.1: Illustration of the estimation scheme of the stability time t_s in the prevalence via SSIS and the spreading time t_m for one realization $i(t)$. The distribution of the number of infected nodes in the metastable state is shown in the right subgraph. The green line represents the average number of infected nodes with time based on 10^6 realizations. The time is measured in units of $1/\delta$.

with different effective infection rate τ. NIMFA approximates the average number of infected nodes in the metastable state for a complete graph K_N with N nodes as $i_s = \lfloor N\left(1-\frac{1}{\tau(N-1)}\right)\rfloor$. When the effective infection rate τ is above the epidemic threshold τ_c, the average hitting time $E[T_{H_i}]$ exhibits two different regimes in the average fraction y of infected nodes as shown in Figure 2.2. In Regime 1, where $y < \frac{i_s}{N}$, the average hitting time $E[T_{H_i}]$ increases exponentially-like as $e^{\kappa y}$, where the rate κ decreases with the effective infection rate τ. In Regime 2, where $y > \frac{i_s}{N}$, the average hitting time $E[T_{H_i}]$ increases faster than an exponential function.

Figure 2.2 suggests that the average hitting time $E[T_{H_n}]$ scales approximately exponentially with the number n of infected nodes around the average number $E[I(t_s)]$ of infected nodes in the metastable state. Assuming that the hitting time T_{H_n} with small variance is correlated to the number n of infected nodes $T_{H_n} \propto e^{\kappa n}$, the spreading time can be regarded as the random variable $T_m(I(t_s)) \approx e^{\kappa I(t_s)+b}$, where the number of the infected nodes $I(t_s)$ is approximately a Gaussian-like random variable [49] with probability density function $\Pr[I(\,t_s) = n] \approx \frac{1}{\tilde{\sigma}\sqrt{2\pi}} \exp\left[-\frac{(n-\tilde{\mu})^2}{2\tilde{\sigma}^2}\right]$. Therefore, we may infer that the pdf of the spreading time is approximately given by

$$f_{T_m}(t) \approx \frac{1}{\kappa t \tilde{\sigma}\sqrt{2\pi}} \exp\left[-\frac{(\frac{1}{\kappa}(\log t - b) - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right] = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\frac{(\log t-\mu)^2}{2\sigma^2}}, \qquad (2.5)$$

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

We first show the spreading time T_m started from one initially infected node in two typical graphs including a complete graph K_{50} and a star $K_{1,49}$ with $N = 50$ nodes.

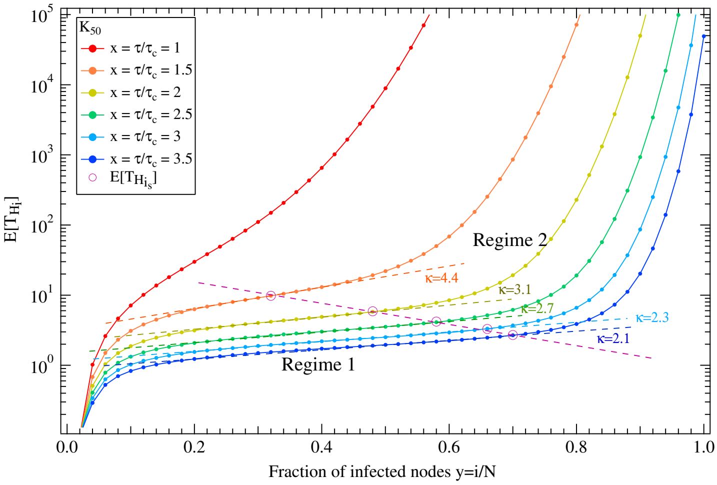


Figure 2.2: The average hitting time $E[T_{H_i}]$ to the state with i infected nodes in the complete graph K_{50} with different effective infection rate τ, given that there is one initially infected node. The average fraction of infected nodes in the metastable state via NIMFA is marked.

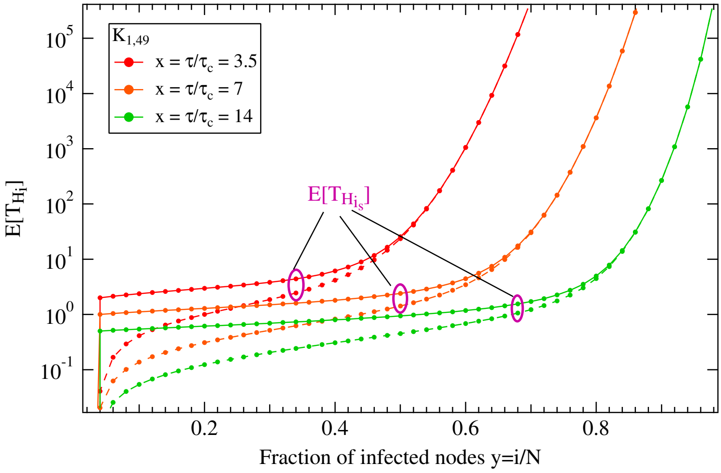


Figure 2.3: The average hitting time $E[T_{H_i}]$ to the state with i infected nodes in the star graph $K_{1,49}$ with different effective infection rate τ, given that there is one initially infected node [47]. The solid line represents the process started from a leaf, and the dash line represents the process started from the center. The average fraction of infected nodes $\frac{i_s}{N}$ in the metastable state via NIMFA is marked.

Figure 2.4 and Figure 2.5 show the spreading time T_m for two values of normalized effective infection rate $x = \tau/\tau_c$ on a log-log scale, based on more than 10^7 realizations. For both graphs, the distribution of the spreading time is fitted by a lognormal pdf (2.5) well around the peak probability, with some deviations in the tail. The positive skewness of the distribution, shown in Figure 2.4 – 2.5, means that the average spreading time $E[T_m]$ is above the mode of the spreading time, which is caused by the rapidly increasing average hitting time $E[T_{H_n}]$ in (2.2), when the number of infected nodes n exceeds the average number i_s of infected nodes in the metastable state. Comparing the distributions with different normalized effective infection rate x in Figure 2.4 and Figure 2.5, the probability of the small value of the spreading time T_m decreases or even disappears with increasing effective infection rate τ.

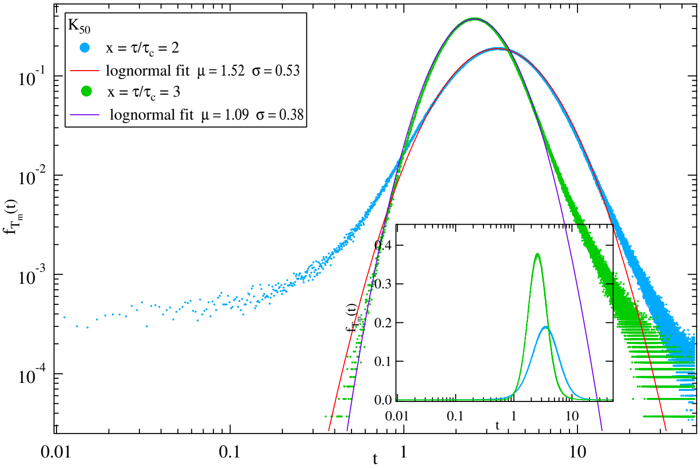


Figure 2.4: The distribution of spreading time T_m in the complete graph K_{50} with the effective infection rate x, which is based on more than 5×10^7 realizations. Both the axes are on log-scale while only x-axis in the subgraph is on log-scale. The skewness of the distribution is 4.8 for $x = 2$ and 11.4 for $x = 3$.

Further, Figure 2.6 – 2.8 show the distributions of the spreading time T_m in an Erdős-Rényi (ER) random graph, a rectangle lattice with $N = 50$ nodes and a BA (Barabási-Albert) power law graph with $N = 1000$ nodes, respectively, where the distribution of the spreading time is influenced by the position of the initially infected spreader and the effective infection rate τ. Taking the lognormal distribution as a reference distribution in the quantile-quantile plots, we find the spreading time also fits the lognormal pdf well when the value of the spreading time is not very large, but deviates in the tail, with a heavier tail than the lognormal distribution. Figure 2.6 presents the distribution of the spreading time T_m in $10^3 \sim 10^7$ realizations for a connected ER random graph $G_{0.2}(50)$. We observe that the deep tails can be reached only when the number of realizations is extremely large (over 10^6 realizations). If the number of realizations is not large enough, the spreading time is restricted around its average without extreme values. Then, the good fit of the distribution by a lognormal pdf may lead to an *incorrect* conclusion that the spreading time is precisely lognormal.

We also observe that the more regular the graph is, the better the distribution of spreading time T_m fits a lognormal pdf. That regularity agrees with the governing rule of a lognormal, as the limit distribution of a sum of the logarithm of random variable that each does not differ much [38]. In the star or the power-law graph, viruses usually need more time to infect one more node with a very small degree. Figure 2.3 for a star

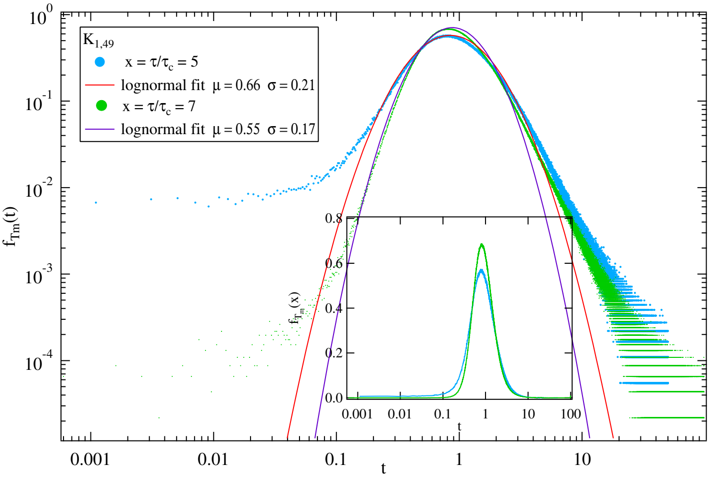


Figure 2.5: The distribution of spreading time T_m in the star $K_{1,49}$ with the effective infection rate x, which in based on more than 10^7 realizations. Both axes are on log-scale while only x-axis in the subgraph is on log-scale. The skewness of the distribution is 11.4 for $x = 5$ and 15.3 for $x = 7$.

graph shows that the function of the hitting time T_{H_i} as the number of infected nodes i increases faster than an exponential around i_s, which may lead to a heavier tail in the distribution of the spreading time, as shown in Figure 2.5. We also mark the stability time t_s via simulation in Figure 2.6 – 2.7, which shows that the stability time t_s lies closely to the tail of the distribution of the spreading time T_m, and is larger than the average spreading time $E[T_m]$.

The infection time is exponentially distributed in the classic Markovian SIS process. More generally, we extend the investigation of the spreading time T_m in a non-Markovian process, which is more common in real-world situations, such as information spread in online social networks and real diseases with incubation periods [50]. We assume

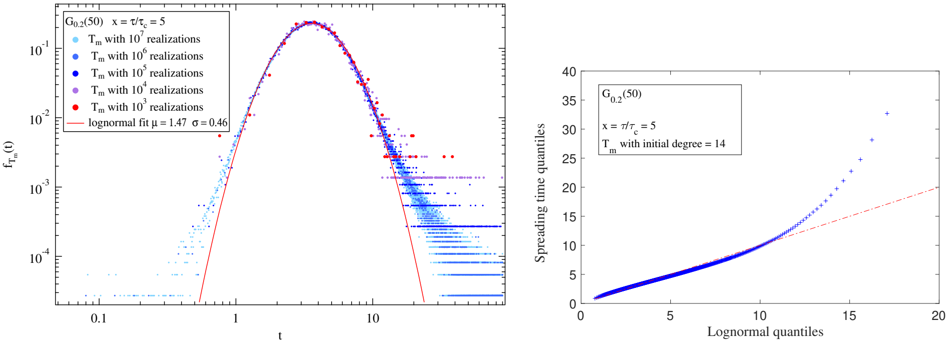


Figure 2.6: The distribution of the spreading time T_m in a connected ER random graph $G_{0.2}(50)$ with 50 nodes started from the initial spreader with $d_{initial} = 4$, $x = \tau/\tau_c = 5$.

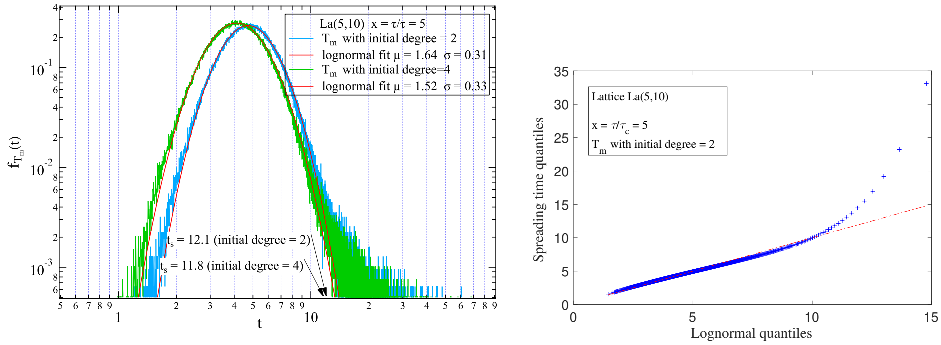


Figure 2.7: The distribution of the spreading time T_m in a grid $La(5,10)$ with 5 × 10 nodes started from the initial node with $d_{initial} = 2$ and $d_{initial} = 4$, $x = \tau/\tau_c = 5$ where $\tau = 1.35$. The histogram is based on 5×10^5 realizations.

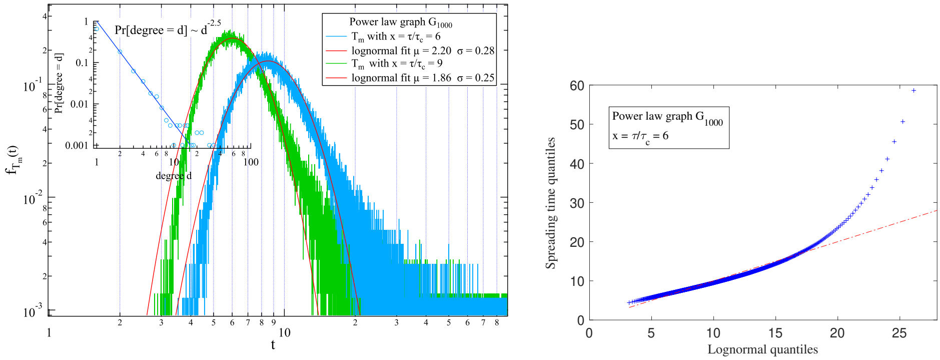


Figure 2.8: The distribution of the spreading time T_m in a power law graph G_{1000} with 1000 nodes starting from one initial node. The histogram is based on 2×10^5 realizations.

that the infection and curing processes are independent in a non-Markovian SIS model, where the curing process is still Poissionian with rate δ, and the infection process at each node infects its neighbors in a time T that is Weibullean, with the pdf

$$f_T(x) = \frac{\alpha}{b}\left(\frac{x}{b}\right)^{\alpha-1} e^{-(x/b)^{\alpha}}. \tag{2.6}$$

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

Figure 2.9 and 2.10 show the distribution of spreading time T_m as a function of the shape parameter α in a complete graph and a star graph. The pdf of the spreading time remains heavy-tailed, and the shape parameter α shifts the mode of the pdf of the spreading time. The tail of the distribution of the spreading time tends to a lognormal pdf better with the increasing shape parameter α in the complete graph.

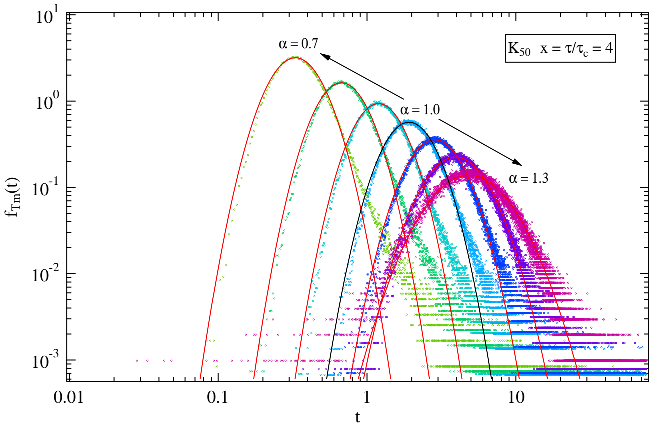


Figure 2.9: The distribution of the spreading time T_m with the different shape parameter α in the complete graph K_{50} with the effective infection rate $x = \tau/\tau_c = 4$. The exponential case ($\alpha = 1$) is indicated in black. The histograms are based on more than 5×10^5 realizations.

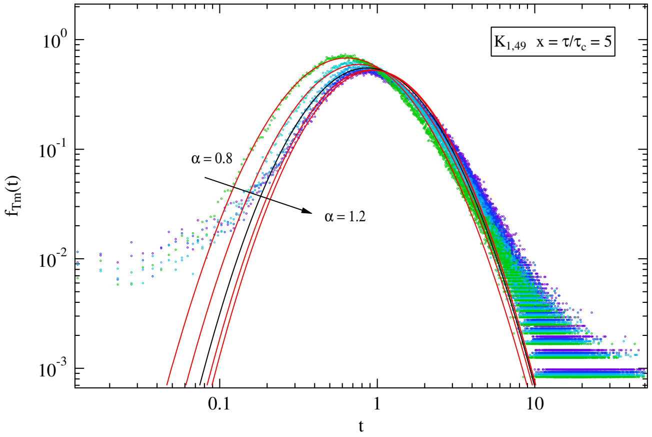


Figure 2.10: The distribution of the spreading time T_m with the different shape parameter α in the star graph $K_{1,49}$ with the effective infection rate $x = \tau/\tau_c = 5$. The exponential case ($\alpha = 1$) is indicated in black. The histograms are based on more than 5×10^5 realizations.

The characteristic times with heavy-tailed distribution in Markovian processes have been observed in a few previous research, such as the inter-record time in the extremal process [51], the time of ruin in the risk model [52] and the first return time of random walks [53]. In this section, we show that the spreading time in the SIS model on a network resembles a lognormal-like distribution with different deep tails, regardless of the process being Markovian or non-Markovian, the network topology and the initially infected node.

2.4. THE AVERAGE SPREADING TIME $E[T_m]$ IN SIS PROCESSES

2.4.1. EFFECT OF THE EFFECTIVE INFECTION RATE ON $E[T_m]$

We study the average spreading time $E[T_m]$ as a function of the effective infection rate τ in a SIS process, started from a same initially infected node. Figure 2.11 and Figure 2.12 illustrate the function of the average spreading time $E[T_m]$ with the effective infection rate τ in a complete graph and a star. The average spreading time $E[T_m]$ is not monotonic with the effective infection rate τ but exhibits a maximum, which means that a stronger virus may not lead to a shorter average spreading time $E[T_m]$.

To better explain the above phenomenon, we define the spreading capacity as $c = \frac{E[I_m]}{E[T_m]}$, which approximately indicates the average number of nodes that can be infected in a time unit in the early state of the spreading. Thus, a higher effective infection rate leads to a smaller reciprocal of the spreading capacity $1/c$, which describes the average time units to infect per node. Meanwhile, the average number of infected nodes $E[I_m]$ in the metastable state increases with the effective infection rate τ in a network when the effective infection rate is above the epidemic threshold τ_c. Therefore, the average spreading time $E[T_m]$, which is represented by $E[T_m] = \frac{E[I_m]}{c}$, is influenced by $E[I_m]$ and the spreading capacity c simultaneously, exhibits the property of non-monotony with the effective infection rate τ. The sub-graphs of Figure 2.11 and Figure 2.12 illustrate the reciprocal of the spreading capacity $1/c$ and the average number of infected node $E[I_m]$ in the metastable state as a function of the effective infection rate τ.

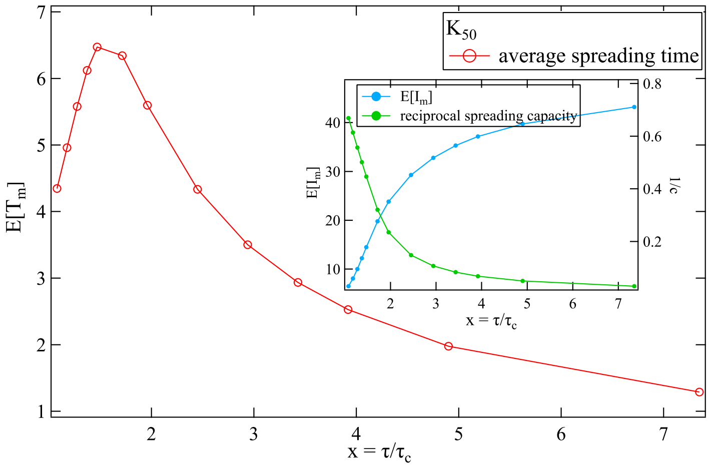


Figure 2.11: The average spreading time $E[T_m]$ as a function of the effective infection rate $x = \tau/\tau_c$ in a complete graph K_{50}. The subgraph illustrates the average number $E[I_m]$ of infected nodes in the metastable state and the the reciprocal of the spreading capacity $1/c$ with the normalized effective infection rate $x = \tau/\tau_c$.

2.4.2. EFFECT OF THE SHAPE PARAMETER α ON $E[T_m]$

We now investigate the effect of the shape parameter α in the Weibull-distributed infection time with pdf (2.6) on the average spreading time $E[T_m]$, where the Markovian

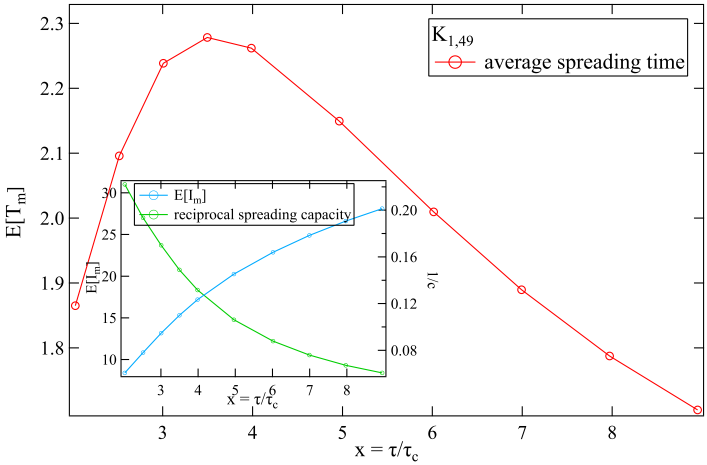


Figure 2.12: The average spreading time $E[T_m]$ with $x = \tau/\tau_c$ in a star graph $K_{1,49}$ with 49 leaves, started from the center of the graph. The subgraph illustrates the average number $E[I_m]$ of infected nodes in the metastable state and the the reciprocal of the spreading capacity $1/c$ with the normalized effective infection rate $x = \tau/\tau_c$.

infection process is a special case with $\alpha = 1$. As discussed in Section 4.1, the average spreading time depends on the spreading capacity c and the average fraction $y(t_s)$ of infected nodes in the metastable state, both of which are influenced by the shape parameter α.

The average number of infection attempts during a recovery time is a physically more general description than the effective infection rate in non-Markovian epidemics [50]. Considering the distribution of the infection attempts over an infectious period of a node, the occurrence of events is not uniformly distributed over an interval when the infection process is non-Markovian. For $\alpha < 1$, the infection events tend to happen earlier than the Poission-distributed events (for $\alpha = 1$) with high probability, while for $\alpha > 1$, the infection events tend to happen later. Therefore, the timing of the infection attempts relative to the curing time of a node influences the epidemics process even for the same average number of expected infection attempts [41]. Physically, the reciprocal of the spreading capacity $1/c$, which describes the average time units to infect per node before the metastable state, also increases for a higher α.

Figure 2.13 shows that the average fraction $y(t_s)$ of infected nodes in the metastable state depends on both the effective infection rate τ and the shape parameter α. Specifically, the average fraction $y(t_s)$ of infected nodes in the metastable state decreases with a higher parameter α for a same effective infection rate τ. Figure 2.14 suggests that $\log(\tau) \sim \frac{\log(Ny(t_s))}{\alpha}$ for the same number $Ny(t_s)$ of infected nodes in the metastable state, which implies that $\tau^\alpha \sim y(t_s)$ in the complete graph when $\tau < 1$. This relation is consistent with the conclusion that the epidemic threshold $\tau_c(\alpha)$ in the non-Markovian SIS epidemics scales as $(\tau_c^{(1)})^{\frac{1}{\alpha}}$, where $\tau_c^{(1)} = \tau_c(1)$ is the epidemic threshold in the Markovian SIS model [40]. As Figure 2.14 shows in the star graph, the Weibull shape factor α barely influences the fraction $y(t_s)$ of infected nodes in the metastable state when the effective infection rate $\tau \geq 1$.

Figure 2.15 shows that, both in the complete graph and the star, the average spreading time $E[T_m]$ does not always increase monotonically with the shape parameter α, but exhibits a maximum when the effective infection rate τ is small. For a higher α, the timing of the infection attempts is postponed while the fraction of infected nodes in the metastable state decreases. These two factors leads to the non-monotonicity of the average spreading time $E[T_m]$ with the shape parameter α, and implies that increasing the parameter α may not shorten the average spreading time $E[T_m]$.

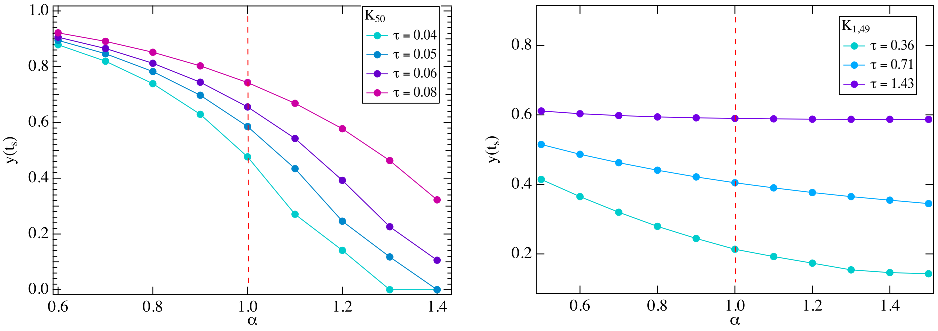


Figure 2.13: The average fraction of infected nodes in the metastable state for the same τ in the non-Markovian SIS process in a complete graph K_{50} and a star graph $K_{1,49}$.

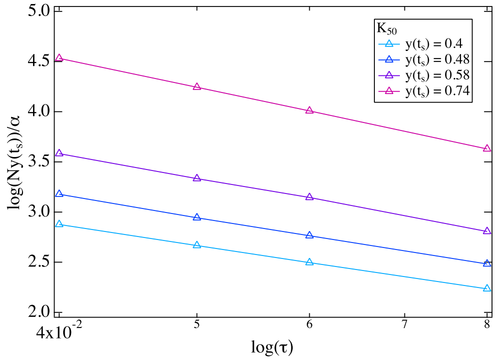


Figure 2.14: The reciprocal of the parameter α as a function of $log(\tau)$ in the complete graph K_{50} for the same fraction of infected nodes in the metastable state.

2.4.3. EFFECT OF THE NETWORK SIZE ON $E[T_m]$

We now investigate the effect of the network size N on the average spreading time $E[T_m]$. Figure 2.16a – 2.16c show the average spreading time $E[T_m]$ starting from one initially infected node as a function of the network size for a complete graph K_N, a star

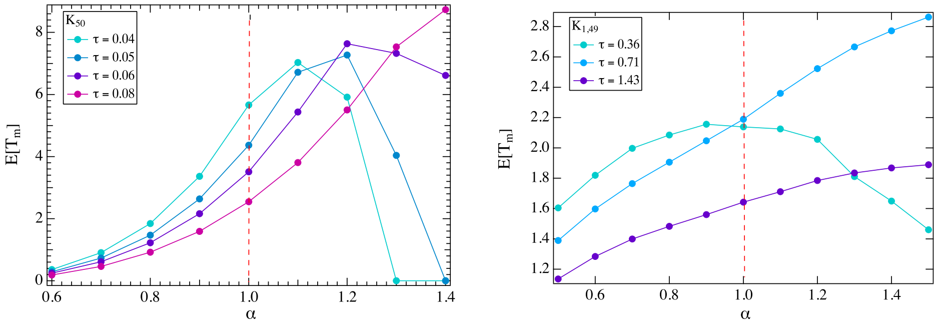


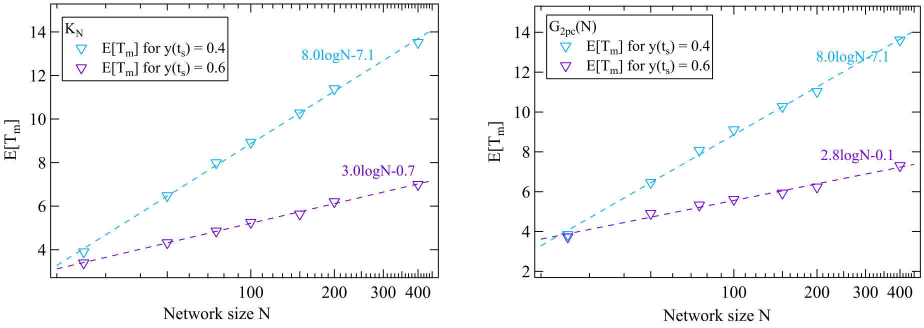
Figure 2.15: The average spreading time $E[T_m]$ as a function of the parameter α for the same effective infection rate τ in the complete graph K_{50} and in the star graph $K_{1,49}$.

$K_{1,N-1}$, and an ER random graph $G_p(N)$. Referring to the average fraction of infected nodes $y^{(1)}(t_s) = 1 - \frac{1}{(N-1)\tau}$ in the metastable state in a complete graph with N nodes via NIMFA [49], we can estimate the effective infection rate $\tau = \frac{1}{(1-y(t_s))(N-1)}$ for a fixed average fraction $y(t_s)$ of infected nodes in the metastable state. Similarly, in an ER graph, the effective infection rate $\tau = \frac{1}{(1-y(t_s))(N-1)p}$ for a fixed fraction $y(t_s)$ of infected nodes in the metastable state is estimated by the NIMFA approximation $y^{(1)}(t_s) = 1 - \frac{1}{(N-1)p\tau}$, where the link probability $p = \frac{2\log N}{N}$. The effective infection rate τ in a star is estimated by the NIMFA approximation [49] that $y^{(1)}(t_s) = \frac{N-\tau^{-2}}{N+1}\left\{\frac{1}{\tau^{-1}+1} + \frac{1}{\tau^{-1}+N}\right\} \approx \frac{\tau}{1+\tau}$ when $N \gg \tau$.

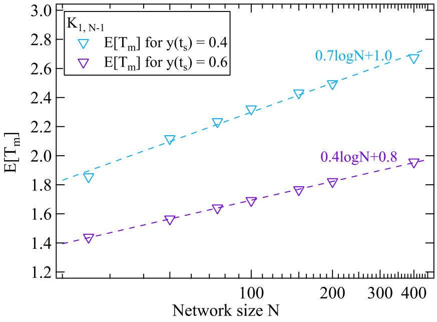
We ignore the curing events and consider a Susceptible-Infected (SI) process in the complete graph. The average time when I_m nodes are infected [38] follows $\sum_{n=1}^{I_m} \frac{1}{\tau n(N-n)}$, where $y(t_s) = \frac{I_m}{N} \approx 1 - \frac{1}{N\tau}$ is fixed. Thus, we obtain

$$E[T_m] \approx \sum_{n=1}^{I_m} \frac{1}{\tau n(N-n)} = \frac{2}{\tau N} \sum_{n=1}^{I_m} \frac{1}{n} \sim 2(1-y(t_s))\log(y(t_s)N), \tag{2.7}$$

which scales logarithmically with the network size N. For an SIS process, Figure 2.16a – 2.16c show that the average spreading time $E[T_m]$ via simulation approximately scales logarithmically as $a\log(N) + b$ for different fractions $y(t_s)$ of infected nodes in the metastable state in a complete graph, an ER random graph and a star. The process needs more time to infect a same fraction of nodes in a network with a larger size. The slope a of the fit is larger for a smaller fraction $y(t_s)$ of infected nodes in the metastable state, which means the average spreading time $E[T_m]$ tends to increase more quickly with the network size N when the fraction $y(t_s)$ of infected nodes in the metastable state is smaller. We observe the similar trend of the average spreading time with the network size N in the complete graph and the ER random graph. Actually, when the link density p in an ER random graph is above the critical link density $p_c = \frac{\log N}{N}$, the graph is already dense and follows similar behaviors as the complete graph [38].



(a) Complete graph K_N (b) ER random graph $G_{2p_c}(N)$



(c) Star graph $K_{1,N-1}$

Figure 2.16: The average spreading time $E[T_m]$ starting from one initially infected node as a function of the network size in the graph.

2.5. CHAPTER SUMMARY

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

We further investigated the properties of the average spreading time. Because the number of infected nodes in the metastable state and the spreading capacity are influenced by the effective infection rate simultaneously, the average spreading time $E[T_m]$ is not necessarily monotonous with the effective infection rate τ but exhibits a maximum, which means that a higher effective infection rate τ may not lead to a shorter average spreading time $E[T_m]$. Similarly, both the fraction of infected nodes in the metastable state and the timing of the infection attempts are influenced simultaneously by the parameter α of the Weilbullean infection times, which leads to non-monotonicity of the average spreading time $E[T_m]$ with the shape parameter α. Finally, we showed that the average spreading time $E[T_m]$ scales logarithmically as a function of the network size N, given that the average fraction $y(t_s)$ of infected nodes in the metastable state is fixed.

3

THE FASTEST SPREADER IN SIS EPIDEMICS ON NETWORKS

Identifying the fastest spreaders in epidemics on a network helps to ensure an efficient spreading. By ranking the average spreading time for different spreaders, we show that the fastest spreader may change with the effective infection rate of a SIS epidemic process, which means that the time-dependent influence of a node is usually strongly coupled to the dynamic process and the underlying network. With increasing effective infection rate, we illustrate that the fastest spreader changes from the node with the largest degree to the node with the shortest flooding time. (The flooding time is the minimum time needed to reach all other nodes if the process is reduced to a flooding process.) Furthermore, by taking the local topology around the spreader and the average flooding time into account, we propose the spreading efficiency as a metric to quantify the efficiency of a spreader and identify the fastest spreader, which is adaptive to different infection rates in general networks.

This chapter is based on the published paper [54].

3.1. INTRODUCTION

IDENTIFYING the most influential initial spreaders in a network constitutes a basic endeavor in network science, which helps to optimize the utility of resources and to ensure an efficient diffusion [55]. Injecting information in the fastest spreaders results in the most efficient spreading performance. The knowledge of the fastest spreader can be applied in direct marking [56] or idea spreading [57], where the resources are limited to start the spreading with a small number of spreaders.

Many topological metrics have been proposed to measure the influence of nodes in networks [58], such as degree, betweenness, closeness [38], eigenvector centrality [59] and the square eigenvector component [60]. Kitsak *et al.* [30] suggest that coreness constitutes a better topological descriptor to identify influential spreaders in epidemics [30]. However, many nodes performing differently in a spreading process may have the same k-core value. Therefore, new metrics based on the existing centrality are proposed to improve the identification of the influential nodes by coreness [61][62]. Considering removing the nodes causing the biggest drop in the energy function, Morone and Makse [63] propose the metric of collective influence through optimal percolation, which performs well in locally tree-like networks. Van Mieghem *et al.* [64] propose that the best conduction node in a resistor network is the minimizer of the diagonal elements of the pseudoinverse matrix $Q^{\dagger}$ of the weighted Laplacian matrix of the graph.

In the SIR (Susceptible-Infected-Removed) model [12], Šikić *et al.* [65] show that the ranking of nodal influences is sensitive to the spreading dynamics, which depends on the infection rate and the curing rate. Measured by the cumulative infection probabilities of nodes, the degree centrality can better identify influential spreaders when the spreading rate is very small. However, the eigenvector centrality performs better when the spreading rate is close to the epidemic threshold [66]. Holme [67] discovers similar results and proposes an exact method to identify the best spreaders for influence maximization (the expected outbreak size) in the SIR model, but the method is only tractable in small graphs. In the SIS (Susceptible-Infected-Susceptible) model, Qu *et al.* [68] unveil that the ranking of nodal metastable infection probability also changes with the effective infection rate.

The "influence" of the spreader in the SIS model is not well defined. In this chapter, we confine ourselves to the *spreading time* $T_m(i)$, defined as the time [35] when the number of infected nodes in the metastable state is first reached, started with one initially infected node i. The spreading time of an epidemic process generally determines the preferred period to take immunization actions to eradicate the spreading [47]. We investigate the average spreading time $E[T_m(i)]$ to identify *the fastest spreader* in an SIS epidemic on a general network.

This chapter is organized as follows. Section 3.2 introduces the spreading time and shows that the average spreading time depends on the topological metrics in an ER random graph. Section 3.3 shows that the fastest spreader changes with the dynamic process in SIS epidemics. Further, we propose the spreading efficiency to identify the fastest spreader. We show the performance in four artificial and real networks in Section 3.4. Finally, we conclude our results in Section 3.5.

3.2. THE SPREADING TIME IN EPIDEMICS ON NETWORKS

We concentrate on the Markovian SIS epidemics [26] on networks, where both the curing and infection processes are Poisson processes. In the SIS epidemics model on a network G with N nodes and L links, the ratio between the infection rate β and the curing rate δ is called the effective infection rate $\tau = \beta/\delta$. The SIS model features a phase transition [39] around the epidemic threshold τ_c. Viruses with an effective infection rate τ above the epidemic threshold τ_c can infect a sizable portion of the population and stay for a long time in the network. A first-order mean-field approximation of the epidemic threshold $\tau_c^{(1)} = 1/\lambda_1$, where λ_1 is the spectral radius of the adjacency matrix A of the network G, was shown to be a lower bound for the epidemic threshold [26]. We denote by $x = \tau/\tau_c^{(1)}$ the normalized effective infection rate.

The spreading time $T_m(i)$ of the Markovian SIS process resembles a lognormal-like distribution with deep tails [35]. The average spreading time $E[T_m(i)]$ approximates the average hitting time when the average fraction y_∞ of infected nodes in the metastable state is reached. Physically, the spreading time $T_m(i)$ describes the spreading velocity in the early stage of the spreading process, which depends on the local topology around the initial spreader i. The analytic expression of the spreading time in a general graph is hard to derive in closed form [47]. Due to the limitation of the analytical methods, an event-driven simulator SSIS for the SIS spreading process based on the Gillespie algorithm is implemented to determine the spreading time [35].

A faster initial spreader speeds up the spreading in the outbreak period and leads to a shorter average spreading time, which measures the efficiency of the spreader. We can identify the fastest nodes by ranking the average spreading time. We first show the effect of the topological properties of the spreader i on the average spreading time $E[T_m(i)]$ in a SIS epidemics on an Erdős-Rényi (ER) random network. Figure 3.1 shows the normalized topological metrics of node i versus the average spreading time $E[T_m(i)]$, which demonstrates that the average spreading time $E[T_m(i)]$ depends on the topological properties of initial spreader i. Specifically, the degree and the closeness of the initial spreader seem to have a similar behavior as the average spreading time in the ER random graph, while the betweenness of the initial spreader has a weaker correlation with the average spreading time. The reciprocal of the diagonal element $(Q^\dagger_{ii})$ of the pseudoinverse matrix $Q^\dagger$ also performs well in ranking the fastest spreaders and behaves similarly as the degree in the ER random graph [64]. Figure 3.1 illustrates that the nodes with the same coreness may occupy a large proportion of the network so that the fastest spreader cannot be identified well by their coreness.

3.3. THE FASTEST SPREADER IN SIS EPIDEMICS

In this section, we further investigate the fastest spreader in the SIS epidemics. The change of the fastest spreader with the effective infection rate τ is presented in an exemplified barbell-like graph. Then, we propose a new metric to identify the fastest spreader.

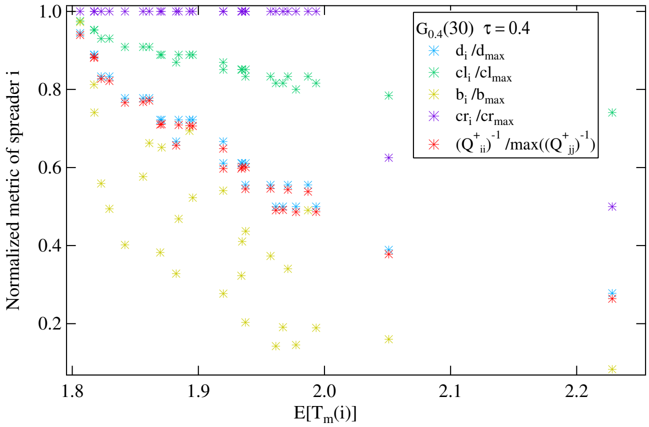
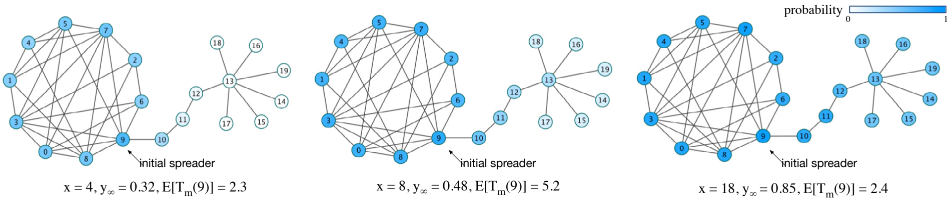


Figure 3.1: The normalized topological metrics of the initial spreader i, e.g., the degree d_i, the betweenness b_i, the closeness cl_i, the coreness cr_i and the reciprocal of the diagonal element $(Q^\dagger_{ii})$ of the pseudoinverse matrix $Q^\dagger$, versus the average spreading time $E[T_m(i)]$ in an connected ER random graph $G_{0.4}(30)$ with $N = 30$ nodes and link density $p = 0.4$.



$x = 4, y_\infty = 0.32, E[T_m(9)] = 2.3$ $x = 8, y_\infty = 0.48, E[T_m(9)] = 5.2$ $x = 18, y_\infty = 0.85, E[T_m(9)] = 2.4$

Figure 3.2: The probability that the nodes is infected at the spreading time for different normalized effective infection rate $x = \tau/\tau_c$. Node 9 is the initially infected spreader. The darkness of the nodes represents the probability. The results is based on 10^5 realizations.

3.3.1. CHANGE OF THE FASTEST SPREADER WITH τ IN A BARBELL-LIKE GRAPH

We generate an asymmetric barbell-like graph G_{20} where a path graph L_2 connects an ER random graph $G_{0.5}(10)$ and a star graph $K_{1,7}$, as shown in Figure 3.2. The barbell-like graph helps us to trace the fastest spreader if the effective infection rate τ changes. Figure 3.2 illustrates the probability that the nodes is infected at the spreading time. Figure 3.2 shows that the infected nodes are usually localized around the initial spreader at the spreading time, e.g., the viruses seldom reach node 14 for a small normalized effective infection rate $x = 4$.

Figure 3.4 exemplifies that the fastest spreader changes with the effective infection rate τ in G_{20}. The fastest spreader changes dramatically from the highest degree node to the lowest degree node with increasing effective infection rate τ. Specifically, we observe three different cases in Figure 3.4. If the effective infection rate τ is relatively small, the fastest spreader tends to be located in the dense part (the ER random subgraph) of the

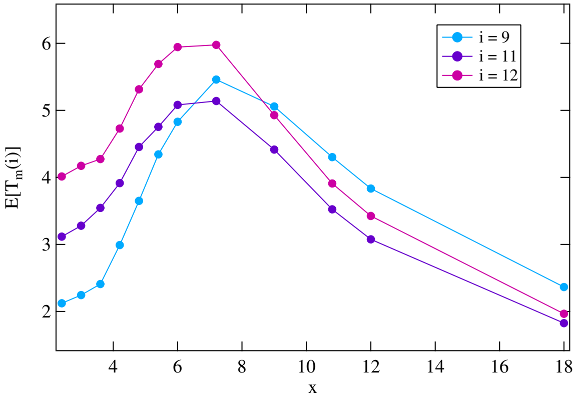


Figure 3.3: The average spreading time $E[T_m(i)]$ as a function of the normalized effective infection rate x in G_{20}, started from node 9, 11 and 12, in the barbell-like graph G_{20}.

network. With the increasing the effective infection rate τ, the fastest spreader transits to nodes with a larger closeness in the path subgraph. At last, the process approximates a flooding process if the effective infection rate τ is large enough. Since the average time to infect all nodes in the star subgraph is larger than that in the ER random subgraph[1], the fastest spreader should be closer to the star subgraph.

In Figure 3.3, the crossings of the average spreading time $E[T_m(i)]$ with the effective infection rate τ for different initial spreaders demonstrate that not only the fastest spreader but also the ranking of spreaders is not fixed for different effective infection rates τ. Therefore, we conclude that the fastest initial spreader in SIS model, only inferred by its location in the underlying graph of the network, cannot be determined. Our finding implies that time-dependent “importance or centrality”of a node is usually strongly coupled to the dynamic process and the underlying graph itself.

3.3.2. A HEURISTIC TOPOLOGICAL METRIC FOR THE FASTEST SPREADER

In this section, we discuss the topological property of the fastest spreader throughout the increase of the effective infection rate τ, i.e., $\tau \downarrow \tau_c$, $\tau > \tau_c$ and very large τ.

CASE: $\tau \downarrow \tau_c$

Invoking the infection probability vector $V(t) = ((v_1(t), v_2(t), \cdots, v_N(t))^T$, we approximate the spreading dynamics in the early stage of the spreading [40] and obtain

$$\frac{dV(t)}{dt} \approx \beta AV(t) - \delta IV(t) \tag{3.1}$$

---

[1] The average time to infect all nodes [69] in an ER random graph $G_p(N)$ is estimated to be $\frac{1}{\beta}\sum_{n=1}^{N-1}\frac{1}{np(N-n)} \sim \frac{2\log(N-1)}{\beta N}$. The average time to infect all nodes in a star graph $K_{1,N}$ from the center is estimated to be the maximum of N exponentially distributed random variables with mean $1/\beta$, which approximates $\sum_{n=1}^{N}\frac{1}{\beta n} \sim \frac{\log N}{\beta}$.

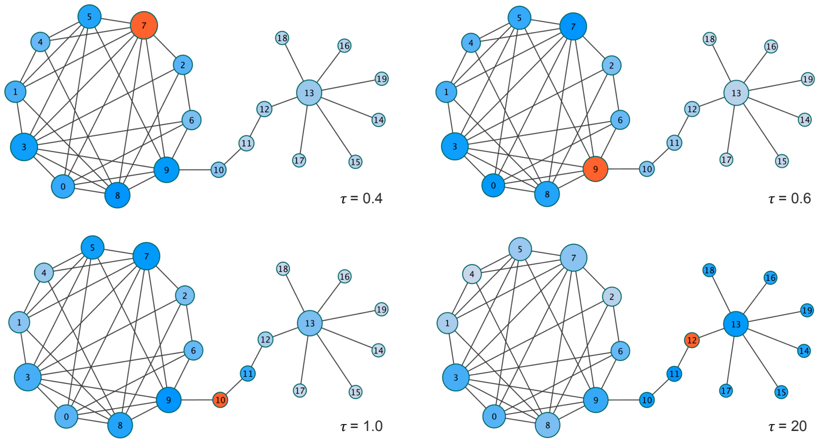


Figure 3.4: Illustration of the changing of the fastest spreader with different τ. The size of the nodes represents the degree, and the darker node represents the faster spreader. The orange node is the fastest initial spreader.

The average fraction $y(t,\tau)$ of infected nodes at the spreading time t_m with τ obeys that

$$y(t_m,\tau) = \frac{1}{N}u^T V(t_m) \approx \frac{1}{N}u^T e^{(\beta A - \delta I)t_m} V(0) \tag{3.2}$$

where $u^T = (1,1,\ldots,1)$. If the effective infection rate $\tau = \frac{\beta}{\delta}$ approaches the first order mean-field approximation of the epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$, only a very small proportion $y(t_m,\tau)$ of nodes will be infected in the metastable state. The spreading time t_m, defined as the the first hitting time when $Ny(t_m,\tau)$ nodes are infected in the Markovian SIS process without extinction [35], is finite. Figure 3.3 also exemplifies that the average spreading time is relatively small if $\tau \downarrow \tau_c^{(1)}$. The matrix $(\beta A - \delta I)t_m$ in (3.2) is dominated by the largest eigenvalue $\delta(\tau\lambda_1 - 1)t_m$, which tends to be 0 if the effective infection rate $\tau \downarrow \tau_c^{(1)}$ (by Perron-Frobenius Theorem [38]). Simplified, invoking the degree vector $d = Au$ and $V(0) = e_i$, we arrive at

$$\begin{aligned}\lim_{\tau \downarrow \tau_c^{(1)}} y(t_m,\tau) &\approx \frac{1}{N}u^T\left(I + (\beta A - \delta I)t_m\right)V(0) \\ &\approx \frac{1}{N}\left(u^T + (\beta d^T - \delta u^T)t_m\right)V(0) \\ &= \frac{1}{N}\left(1 - t_m\delta + t_m\beta d_i\right)\end{aligned} \tag{3.3}$$

Relation (3.3) exhibits that the degree of the spreader dominates the spreading time t_m for the unaltered rates β, δ and their corresponding $y(t_m) = y_\infty$. This result is different from the result that the eigenvector of the adjacent matrix A belonging to the largest eigenvalue determines the infection probability vector in the metastable state [38]. We here exemplify an extreme case: if the effective infection rate τ approaches τ_c, and there is only one infected node in the metastable stable, i.e., $y(t_m) = \frac{1}{N}$, the spreading time t_m equals the minimum time when any one of the neighbors of the spreader i is infected. Then, the average spreading time $E[T_m]$ is the minimum of the d_i exponential distributed random variables with a mean $1/\beta$, where d_i is the degree of the spreader i. Thus, the average spreading time follows $E[T_m] = \frac{1}{\beta d_i}$, which is determined by the degree of the initial spreader.

CASE: INCREASING τ

We then investigate the case for the increasing effective infection rate τ. Inspired by the illustration in Section 3.3.1, we postulate that the fastest spreader depends on the *local* topology around itself, i.e., the number of nodes and the connectivity of nodes around the spreader. We first consider the number of nodes around the initial spreader and regard that the efficiency of the initial spreader is related to the expansion [38] of the subgraph centered at the spreader. Specifically, assuming that the hop count h is the farthest distance from the initial spreader i that the viruses can reach before the spreading time, the expansion of the subgraph is the number of nodes $|C_i(h)|$ within h hops from the initial spreader i.

We then consider the connectivity of the nodes around the initial spreader. An epidemic behaves like a continuous time Markov branching process in the early stage [70]. For a branching process, we obtain that the number of infected nodes follows

$$Ny(t) \approx u^T e^{\beta A t} V(0) \leq e^{\beta \lambda_1 t} Ny(0) \tag{3.4}$$

which implies that the lower bound of the time to infect $Ny(t)$ nodes around the initial spreader follows that $t \geq \frac{\log(Ny(t))}{\beta \lambda_1}$. Inspired by (3.4), we propose $\frac{\lambda_i(h)}{\log|C_i(h)|}$ as an indication of the connectivity of the local topology around the spreader i for a fixed infection rate β, where $\lambda_i(h)$ is the largest eigenvalue of the subgraph within h hops around the initial spreader i. A larger $\frac{\lambda_i(h)}{\log|C_i(h)|}$ implies a higher connectivity that leads to a faster spreading in the local network within h hops.

Considering the above two factors including the expansion $|C_i(h)|$ of the subgraph and the connectivity indication $\frac{\lambda_i(h)}{\log|C_i(h)|}$ within the subgraph, we propose the spreading efficiency as a new metric to measure the efficiency of the initial spreader in the SIS model. The spreading efficiency of node i is defined as

$$\mathcal{E}_i = \frac{\lambda_i(h)}{\log|C_i(h)|} |C_i(h)| \tag{3.5}$$

In case that the sub-graph expansion $|C_i(h)|$ of the initial spreaders are the same, a larger sub-eigenvalue $\lambda_i(h)$ leads a higher spreading efficiency in the subgraph due to a higher connectivity of nodes.

The hop count h describes the average farthest distance of the infected nodes from the spreader at the spreading time for the effective infection rate τ, which is difficult to be determined precisely in a general network. Morone and Makse [63] identify the influential spreaders by the Ball (subgraph) centered at the spreader, where the optimal radius of the Ball is 3 or 4. The optimal hop $h = f(\tau)$ in our method is more flexible, which is a function of the effective infection rate τ. We hereby proceed with an approximation. First, the average fraction of infected nodes y_∞ in the metastable state can be estimated by the NIMFA approach for a determined τ. The number N_C of nodes in a branch process follows $N_C \approx \frac{\mu^{H+1}-1}{\mu-1}$, where H is the largest hop count from the root and $\mu = E[D]-1$ is the mean degree minus 1 in this graph [38]. In that case, we have the largest hop count $H \approx \frac{\log(N_C(\mu-1)+1)}{\log\mu} - 1 \approx \frac{\log N_C}{\log\mu}$ if $N_C \gg \mu$. Invoking the fact that a spreading process approximates a branching process in the early stage, we can estimate the hop count h in a sparse and large graph by

$$h = \left\lceil \frac{\log N y_\infty}{\log\mu} \right\rceil \tag{3.6}$$

CASE: LARGE τ

With the increase of the effective infection rate τ and the average fraction y_∞ of infected nodes in the metastable state, the nodes that need relatively more time to be reached gradually dominate the spreading time. Thus, the fastest spreader could be closer to the sparser subgraph of the network.

Finally, if the effective infection rate τ is large enough, the SIS process is reduced to be a flooding process [69]. The average flooding time $E[T_N(i)]$ of an initial spreader i is the average minimum time for the virus to reach all other nodes in a flooding process. Therefore, we could regard the reciprocal of the average flooding time $\phi_i = \frac{1}{E[T_N(i)]}$ determines the fastest spreader if τ is very large.

Assuming that $\lambda_1(0) = 1$ and $|C_i(0)| = d_i$, the spreading efficiency in (3.5) with $h < 1$ follows the same rank as the degree d_i. In summary, we simplify and propose the overall metric "spreading efficiency" to identify the fastest initial spreader in an SIS epidemics as

$$\mathcal{E}_i = \begin{cases} \dfrac{\lambda_1(h)|C_i(h)|}{\log|C_i(h)|} & y_\infty \le y_\infty^* \\ \phi_i & y_\infty > y_\infty^* \end{cases} \tag{3.7}$$

where y_∞^* is a prescribed parameter indicating that the process approximates a flooding process if $y_\infty > y_\infty^*$. We set $y_\infty^* = 0.8$ in this chapter for the simulation.

Figure 3.5 shows the Kendall rank correlation coefficient κ between the average spreading time and the above discussed metrics, including the degree, the spreading efficiency in (3.5) and the reciprocal of the average flooding time ϕ_i via Monte-Carlo estimation. If the effective infection τ is close to the epidemics threshold $\tau_c \approx 0.17$, the degree centrality could be a better metric. We then observe that the best hop count h increases with the effective infection rate τ, and the spreading efficiency $\mathcal{E}_i$ with the proposed hop count h in (3.6) can lead to the maximum correlation coefficient κ in a wide range of τ. At last, the reciprocal of the average flooding time shows the advantage when τ is large enough.

3.4. NUMERICAL RESULTS

We evaluate the performance by identifying the ranking of the fastest initial spreaders in four, artificial and real, networks with different sizes and topologies: co-appearances of characters in Les Misérables [71], small world citation network (SmallWCitation) [72], the artificial barbell network G_{20} and Co-authorship network of scientists (Net-Science)

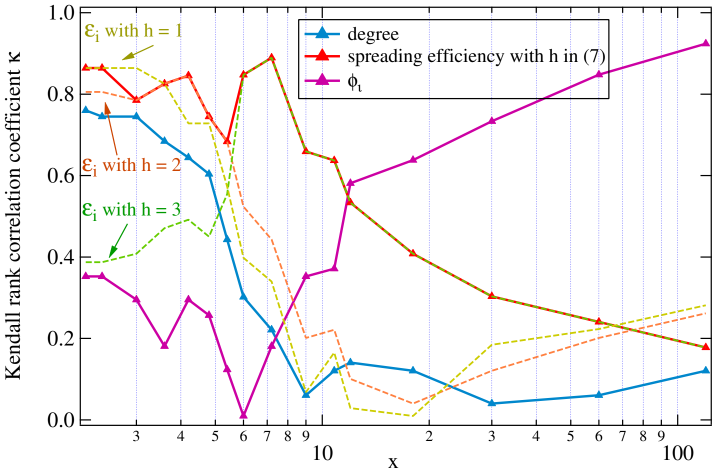


Figure 3.5: The Kendall rank correlation coefficient κ between the average spreading time $E[T_m(i)]$ and the metrics: the degree d_i, the spreading efficiency $\mathcal{E}_i$ and the reciprocal of the average flooding time ϕ_i in the barbell graph G_{20}.

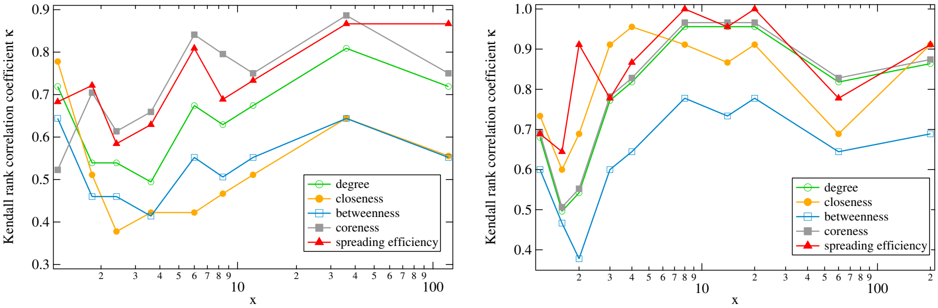
[73]. Table 3.1 shows some properties of the giant component of the four networks including the number of nodes N, the number of links L, the diameter ρ, the largest eigenvalue λ_1, the clustering coefficient C_G, the Pearson degree correlation coefficient ρ_D.

	N	L	ρ	λ_1	C_G	ρ_D
Les Misérables	77	254	5	12.01	0.57	-0.17
SmallWCitation	233	994	4	20.96	0.56	-0.30
Barbell G_{20}	20	39	7	6.14	0.33	-0.11
NetScience	379	914	17	10.38	0.74	-0.08

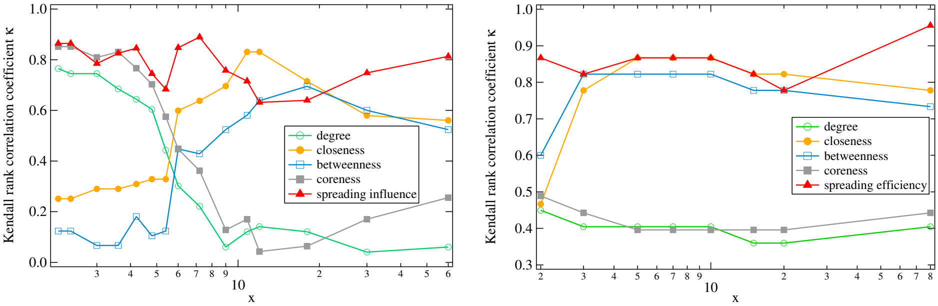
Table 3.1: The topological properties of the giant component of the four experimental networks

We extract the giant component of the above network and select 10 nodes randomly in each network. In each implementation, only one of the selected nodes is infected initially, and then the virus spreads in the network according to the Markovian SIS model. After obtaining the average spreading time via SSIS started from different initial spreaders, we compare the Kendall rank correlation coefficients κ between the average spreading time and some other metrics including degree, closeness, betweenness, coreness and the proposed spreading efficiency in (3.7). Physically, the identification of the fastest spreaders in a flooding process is a 1-center problem [74] in a graph, where the weights of links in the graph are exponentially distributed random variables with mean $\frac{1}{\beta}$. Thus, we estimate the average flooding time $E[T_N(i)]$ by Monte-Carlo approach and the efficiency shortest path algorithm [75].

Figure 3.6 shows the performance of the several centrality metrics for ranking the



(a) Characters network in Les Misérables (b) SmallW citation network



(c) Barbell network G_{20} (d) NetScience network

Figure 3.6: The Kendall rank correlation coefficient κ between the average spreading time $E[T_m(i)]$ and the metrics including degree d_i, closeness cl_i, betweenness b_i coreness cr_i and the spreading efficiency $\mathcal{E}_i$ in four networks.

fastest initial spreader in four networks. We observe that, in the networks with a small diameter (e.g., Les Misérables and SmallW citation network), the spreading efficiency performs similarly with the coreness, both of which are better than other centrality metrics. In addition, the spreading efficiency shows its advantage over the coreness if the effective infection rate τ is relatively large because the reciprocal of the average flooding time determines the fastest spreader in that case.

However, the degree and coreness show the vulnerability in the community networks with a large diameter (e.g., Barbell and NetScience network). Meanwhile, the closeness becomes a better metric, which considers the average length of the path between the spreader and all other nodes. Especially, in the Barbell G_{20}, we observe the changing of the performance of the centrality metrics with the increasing effective infection rate τ. When τ is small, the degree and the coreness perform better, but the closeness and the betweenness become better if τ is large enough, which further convinces us that a single existing centrality metric fails to identify the fastest spreader in the SIS model. The results suggest that, in the real world, the viruses or information may spread more efficiency starting from the spreader with a large degree within the community for a small τ, but it is better to choose the spreader with a high closeness for a large τ.

In summary, we can observe that the proposed spreading efficiency performs better than the compared topological metrics in general, which is adaptive to different topologies and different dynamic process. We find that the accuracy of the spreading efficiency drops a little around the effective infection rate corresponding to the transition parameter y_∞^*. We also expect a better transition method and a better estimation of hop $h = f(\tau)$ that can improve the performance.

3.5. CHAPTER SUMMARY

We investigated the properties of the fastest initial spreader with the shortest average spreading time in the SIS model. We showed that the fastest spreader changes from the node with the largest degree to the node with the shortest flooding time for the increasing effective infection rate, which implies that the fastest spreader is coupled to not only the underlying graph but also the dynamic process.

By considering the expansion and the largest eigenvalue of the subgraph around the spreader, we proposed the spreading efficiency as a metric to rank the fastest spreaders. The spreading efficiency depends on the effective infection rate τ, and reduces to the reciprocal of the flooding time for a large τ. The simulation results on several real-world networks show that the spreading efficiency can better rank the fastest spreaders than some existing topological metrics including degree, closeness, betweenness, and coreness, in different topologies and dynamic processes.

4

OPTIMAL INDUCED SPREADING OF SIS EPIDEMICS IN NETWORKS

Induced spreading aims to maximize the infection probabilities of some target nodes by adjusting the nodal infection rates, which can be applied in biochemical and information spreading. We assume that the adjustment of the nodal infection rates has an associated cost and formulate the induced spreading for SIS epidemics in networks as an optimization problem under a constraint on total cost. We address and solve both a static model and a dynamic model for the optimization of the induced SIS spreading. By numerical results in some artificial and real networks, we investigate the effect of the network topology on the optimal induced strategy with a quadratic cost function. In the static method, the infection rate increment on each node is coupled to both the degree and the average hops to the targets. In the dynamic method, we show that the effective resistance could be a good metric to indicate the minimum total cost for targeting a single node. We also illustrate that the minimum total cost increases much less sharply with the increasing fraction of targets in the SIS model than in linear control systems.

This chapter is based on the published paper [76].

4.1. INTRODUCTION

SINCE the earliest account of mathematical modeling of diseases was proposed by Daniel Bernoulli in 1766, epidemic models help us to better understand dynamics of spreading processes [12]. Spreading in networks can describe many physical phenomena and human activities, such as information spreading on social networks, biological diseases [26] and computer viruses on cyber-physical networks [77]. Some previous research investigated the strategies to eliminate or control the spreading of viruses as quickly as possible [78]. The optimal epidemics control problem for mean-field models can be converted into a spectral control problem, e.g., how to decrease the spectral radius of a graph [28] more efficiently by some strategies (e.g., link removals)? Preciado *et al.* [29] propose the budget-constrained allocation problem and present a solution framework based on Geometric Programming to control the epidemics by adjusting the infection rates and the curing rates of nodes.

Although the existing work on controlling the spreading has presented some useful proposals and frameworks on a microscopic level, the previous research focuses on the performance of the whole network. Induced spreading or targeted spreading is a more general task, which aims to maximize (minimize) the infection probability of some specific nodes instead of all nodes in the network. The induced spreading problem is first introduced by Sun *et al.* [79] for identifying the single best spreader if the spreading only aims to cover a specific group of nodes. The induced spreading problem is inspired by many real applications. In the biochemical application, induced spreading can be applied for targeting biochemical cascades to treat cancer [80]. We prefer to guide the drug to reach affected areas effectively with a minimum dose to reduce side effects. In the context of information spreading, some advertisements (e.g., cigarettes) on the Internet should reach as much as possible to the potential customs; and a strategy for active cyber defense [81] based on spreading patches for targeting the infected computers should be designed.

Notwithstanding the importance of the induced spreading, the methods and properties of the induced spreading have been considered only in a few works [79] [82]. In this chapter, we focus on the Susceptible-Infected-Susceptible (SIS) model and specialize in the optimal induced spreading problem in networks. We address two optimization models: the static and the dynamic. The static optimization aims to maximize the sum of the steady-state infection probabilities of the target nodes, under a constraint on the nodal infection rates. In the dynamic optimization, we model the induced spreading as an optimal control problem for maximizing the cumulative infection probabilities of the target nodes in a time interval, where the time-dependent nodal infection rates are the control variables. For the dynamic optimization, Lokhov and Saad [82] propose a framework for maximizing impact in spreading processes based a message-passing [83] under the assumption of a locally tree-like network. Instead, we formulate both optimization problems based on the heterogeneous NIMFA model [84][85] for better investigating the effect of topology on the induced spreading behaviors.

We solve the static optimization problem by the Differential Evolution algorithm [86], and solve the dynamic optimization based on the optimality system. Further, we investigate the impact of the topological properties of the target nodes on the induced strategy by numerical results in some artificial and real networks. We explore the behaviors of both the static and dynamic induced spreading, as well as compare the performance between these two optimization models.

This chapter is organized as follows. Section 4.2 briefly introduces the SIS model in networks and the heterogeneous NIMFA model. We propose and solve the static optimization for the induced spreading in Section 4.3, and address the dynamic optimization for the induced spreading in Section 4.4. In Section 4.5, we explore some properties of the induced strategy by numerical results. We introduce the related work in Section 4.6 and conclude this chapter in Section 4.7.

4.2. PRELIMINARIES AND MODEL

4.2.1. SIS MODEL IN NETWORKS

We first recall the SIS model. The SIS model is an epidemic model where each infected item can be cured, and becomes susceptible again after recovering from the disease [38]. We define a Bernoulli random variable $X_i(t) \in \{0,1\}$ as the state of a node i at time t, with $X_i(t) = 0$ for the healthy state and $X_i(t) = 1$ for the infected state. The network G with N nodes and L links is represented by an adjacent matrix A, where $a_{ij} = 1$ if there is a link between node i and node j, otherwise $a_{ij} = 0$. We denote by $\mathcal{N} = \{1,2,\dots,N\}$ the set of nodes in the network.

In Markovian Susceptible-Infected-Susceptible (SIS) epidemics, both the curing and infection processes are Poisson processes [38]. Since X_i is a Bernoulli random variable, it holds that $E[X_i(t)] = \Pr[X_i(t) = 1]$, and the exact SIS governing equation for node i equals

$$\frac{dE[X_i(t)]}{dt} = E\left[-\delta X_i(t) + \beta(1 - X_i(t)) \sum_{k=1}^{N} a_{ki} X_k(t)\right] \tag{4.1}$$

The ratio between the infection rate β and the curing rate δ is called the effective infection rate $\tau = \beta/\delta$. The SIS model features a phase transition [39][40] around the epidemic threshold τ_c. Viruses with an effective infection rate τ above the epidemic threshold τ_c can infect a sizeable portion of the population on average and stay for a long time in the network. This long period is called the metastable state [35].

4.2.2. HETEROGENEOUS NIMFA MODEL

In the N-Intertwined Mean-Field Approximation (NIMFA) [26], the infection probability v_i of node i that approximates the exact $E[X_i]$ is given by the following first-order nonlinear ordinary differential equation:

$$\frac{dv_i(t)}{dt} = \beta \sum_{j=1}^{N} a_{ij} v_j(t) - v_i(t)\Big(\beta \sum_{j=1}^{N} a_{ij} v_j(t) + \delta\Big). \tag{4.2}$$

A first-order mean-field approximation of the epidemic threshold $\tau_c^{(1)} = 1/\lambda_1(A)$, where $\lambda_1(A)$ is the spectral radius of the adjacency matrix A, was shown to be a lower bound, i.e., $\tau_c^{(1)} < \tau_c$, for the epidemic threshold [26][40].

Heterogeneous infection is more realistic than the assumption of homogeneity in real-world spreading processes. For example, the transmission capacity per link in a data communication network can be different. In social networks, people who are keen on the social activities could spread a rumor more efficiently. The individual behavior leads to a difference in the infection rate β_i and the curing rate δ_i. The heterogeneous NIMFA model [84][85] with the time-dependent infection rate $\beta_i(t)$ and the curing rate $\delta_i(t)$ of node i is described by

$$\frac{dv_i(t)}{dt} = (1 - v_i(t)) \sum_{j=1}^{N} a_{ij} \beta_j(t) v_j(t) - \delta_i(t) v_i(t) \tag{4.3}$$

In this chapter, we focus on a method to adjusting the nodal infection rates $\beta_i(t)$ for the induced spreading. For simplicity and without lack of generality, we normalize the curing rate by $\delta_i(t) = 1$ at any time for all nodes. Hence, the infection rate $\beta_i(t)$ equals the effective infection rate $\tau_i(t)$.

4.3. STATIC OPTIMIZATION FOR INDUCED SPREADING

4.3.1. PROBLEM STATEMENTS

The SIS process can stay in the metastable state for a much longer time compared to the transient period if the effective infection rate τ is above the epidemic threshold τ_c. Induced spreading for a long term refers to the static optimization of the steady-state infection probability vector $\mathbf{v}_\infty = (v_{1\infty}, v_{2\infty}, \ldots, v_{N\infty})$ by adjusting the infection rates β_j for some nodes $j \in \mathcal{N}$. Static optimization is time-independent and has the advantage of operational simplicity. Specifically, the optimization problem aims to maximize the total steady-state infection probabilities of the nodes in the target set $\mathcal{S}$,

$$\max_{\mathbf{v}_\infty, \boldsymbol{\Delta\beta}} \quad J = \sum_{i \in S} v_{i\infty} \tag{4.4}$$

subject to the steady-state NIMFA equation with the infection rate $\beta_j = \hat{\beta} + \Delta\beta_j$ of node j,

$$(1 - v_{i\infty}) \sum_{j=1}^{N} a_{ij} (\hat{\beta} + \Delta\beta_j) v_{j\infty} - v_{i\infty} = 0, \quad i, j \in \mathcal{N} \tag{4.5}$$

where $\hat{\beta}$ is the original infection rate for all nodes, and the rate increment vector $\boldsymbol{\Delta\beta} = (\Delta\beta_1, \Delta\beta_2, \ldots, \Delta\beta_N)$ are the control variables. Also, we have the constraint on the total cost budget

$$\sum_{i=1}^{N} g(\Delta\beta_i) \leq C, \quad i \in \mathcal{N} \tag{4.6}$$

where $g(\Delta\beta_i)$ is a convex function of the infection rate increment $\Delta\beta_i$, and C denotes a prescribed positive constant for the cost budget. Before solving this problem, we have Lemma 4.1 and Theorem 4.1 as follow:

Lemma 4.1 *The steady-state infection probability $v_{i\infty}$ of any node i in the graph G_N monotonically increases with the infection rate β_j of any node j.*

Proof: See Appendix B.1. ☐

Theorem 4.1 *The steady-state infection probability $v_{i\infty}$ of node i is not always concave with respect to the infection rate β_j for $i \neq j$.*

Proof: See Appendix B.1. ☐

Since the diagonal element $\frac{\partial^2 v_{i\infty}}{\partial \beta_j^2}$ in the Hessian matrix of the steady-state infection probability $\mathbf{v}_\infty$ with respect to the infection rate increment $\boldsymbol{\Delta\beta}$ could be positive according to Theorem 4.1, the Hessian matrix is not negative semi-definite. Thus, we conclude that the static optimization for induced SIS spreading is not a convex program, which cannot be solved by a simple method.

4.3.2. GLOBAL OPTIMIZATION BY DIFFERENTIAL EVOLUTION

We further simplify the NIMFA constraint (4.5) by expressing $v_{i\infty}$ explicitly, and only dependent on the infection rates $\boldsymbol{\beta}$, but not on any other $v_{j\infty}$ for $j \neq i$. For any effective infection rate $\tau \geq 0$, the nonzero steady-state infection probability $v_{i\infty}$ of any node i in the NIMFA can expressed as a continued fraction [26][38]. We define the k-level infection probability for node i as $v_{i\infty}^{(k)} = 1 - w_i(k)$, where the k-th convergent is $w_i(k) = \frac{1}{1+\sum_{j=1}^{N} a_{ij}\beta_j - \sum_{j=1}^{N} a_{ij}\beta_j w_j(k-1)}$ with starting value $w_i(0) = 0$ and $\lim_{k\to\infty} w_i(k) = 1 - v_{i\infty}$.

If the level k is large enough [26], the infection probability $v_{i\infty}^{(k)}$ is sufficiently close to $v_{i\infty}$. By approximating the infection probability $v_{i\infty}$ by $v_{i\infty}^{(k)}$, the optimization problem can be reduced to $\max_{\boldsymbol{\Delta\beta}} \sum\limits_{i\in S} v_{i\infty}^{(k)}$ subject to the cost constraint (4.6). We further reduce the inequality constraint $\sum_{i=1}^{N} g(\Delta\beta_i) - C \leq 0$ to a penalty term by converting the objective to

$$\min_{\boldsymbol{\Delta\beta}} \quad -\sum_{i\in\mathcal{S}} v_i^{(k)} + \zeta \max\{0, \sum_{i=1}^{N} g(\Delta\beta_i) - C\} \tag{4.7}$$

where ζ is the penalty parameter. If the constraint is violated during the optimization process, the penalty term feeds the deviation to the objective function and draws the solution to the feasible region.

Since Problem (4.7) is multi-dimensional and nonlinear, we propose an approach based on the Differential Evolution (DE) algorithm to solve the constrained optimization problem. Differential evolution can approximate the global optima of a nonlinear program, which was proposed by Storn and Price [86]. The Differential Evolution algorithm resembles other traditional evolution algorithms like genetic algorithms (GA), and can represent the solution domain of $\boldsymbol{\Delta\beta}$ by real numbers. The Differential Evolution algorithm has the advantage of implementation simplicity over other non-genetic global optimization algorithms, e.g., the performance of simulated annealing algorithm is sensitive to the cooling rate and the initial solution. Also, the Differential Evolution algorithm [87][88] is usually more efficient and accurate than several other optimization methods, e.g., simulated annealing and genetic algorithms. The proposed Differential Evolution method is based on population generation, mutation, crossover, and selection. The implementation is presented in Algorithm 2.1 (see Appendix B.2).

4.4. DYNAMIC OPTIMIZATION FOR INDUCED SPREADING

4.4.1. PROBLEM STATEMENTS

Dynamic optimization is a more general and more flexible method for the induced spreading. The infection rate increment $\mathbf{\Delta\beta}(t) = (\Delta\beta_1(t), \Delta\beta_2(t), \ldots, \Delta\beta_N(t))$ is time dependent in the dynamic optimization, which is different from the static optimization. Specifically, the dynamic optimization for the induced spreading aims to maximize the total cumulative infection probability $v_i(t)$ of the nodes in the target set $i \in \mathcal{S}$ in the time interval $[t_0, t_f]$, i.e.,

$$\max_{\mathbf{\Delta\beta}(t)} \quad J = \int_{t_0}^{t_f} \sum_{i \in \mathcal{S}} v_i(t)\,dt \tag{4.8}$$

subject to the NIMFA equation

$$\frac{dv_i(t)}{dt} = (1 - v_i(t)) \sum_{j=1}^{N} a_{ij}(\hat{\beta} + \Delta\beta_j(t)) v_j(t) - v_i(t), \quad i, j \in \mathcal{N} \tag{4.9}$$

and the constraint on the cumulative cost in the time interval $[t_0, t_f]$

$$\int_{t_0}^{t_f} \sum_{i \in \mathcal{N}} g(\Delta\beta_i(t))\,dt \le C, \quad i \in \mathcal{N} \tag{4.10}$$

where C is a prescribed cost budget. The dynamic optimization for induced SIS spreading is a control-affine nonlinear model [89] with the integral constraint (4.10), which cannot be solved by the standard methods for Linear Quadratic Regulator (LQR) problems.

4.4.2. THE OPTIMAL SOLUTION

We first introduce an additional control variable z and rewrite the cost constraint (4.10) as

$$\frac{dz(t)}{dt} = \sum_{i \in \mathcal{N}} g(\Delta\beta_i(t)) \tag{4.11}$$

with $z(t_0) = 0$ and $z(t_f) = C$, where t_0 is the initial time and t_f is the final time. Then, the corresponding Hamiltonian is

$$H(\mathbf{v}, \mathbf{\Delta\beta}, \boldsymbol{\theta}, \mu) = \sum_{m \in \mathcal{S}} v_m + \mu \sum_{i \in \mathcal{N}} g(\Delta\beta_i(t)) + \sum_{i=1}^{N} \theta_i \left[(1 - v_i) \sum_{j=1}^{N} a_{ij}(\hat{\beta}_j + \Delta\beta_j) v_j - \delta v_i \right] \tag{4.12}$$

where the parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)$ and μ are undetermined. Next, we present the optimality conditions for the dynamic optimization.

Theorem 4.2 *Suppose* $\mathbf{\Delta\beta}^*(t)$ *is an optimal control for the problem, and* $\mathbf{v}^*(t)$ *is the optimal solution with* $\mathbf{\Delta\beta}(t) = \mathbf{\Delta\beta}^*(t)$*. Then, there exist functions* $\boldsymbol{\theta}^*(t)$ *and* $\mu^*(t)$*, such that*

$$\begin{cases} \dfrac{d\theta_i^*}{dt} = \chi_i + \theta_i^* \left(\sum_{j=1}^{N} a_{ij} \hat{\beta}_j v_j^* + 1 \right) - \hat{\beta}_i \sum_{j=1}^{N} a_{ij}(1 - v_j^*)\theta_j^*, \quad i \in \mathcal{N} \\ \dfrac{d\mu^*}{dt} = 0 \end{cases} \tag{4.13}$$

with the terminal (transversality) conditions $\mu^*(t_f) = 0$ *and* $\theta_i^*(t_f) = 0$ *for* $i = 1,2,\dots,N$*, where* $\chi_i = -1$ *for* $i \in \mathcal{S}$*, and* $\chi_i = 0$ *for* $i \notin \mathcal{S}$*.*

Furthermore, the optimal control variable $\boldsymbol{\Delta\beta}^*(t)$ *obeys*

$$\frac{dg(\Delta\beta_i^*)}{d\Delta\beta_i^*} = -\frac{1}{\mu} v_i^* \sum_{j=1}^{N} a_{ij}(1-v_j^*)\theta_j^*, \quad i \in \mathcal{N} \tag{4.14}$$

Proof: According to the Pontryagin Maximum Principle [90], we can obtain the costate equations as

$$\begin{cases} \dfrac{d\theta_i^*}{dt} = -\dfrac{\partial H(\mathbf{v}^*,\boldsymbol{\Delta\beta}^*,\boldsymbol{\theta}^*,\mu^*)}{\partial v_i} \\ \dfrac{d\mu^*}{dt} = -\dfrac{\partial H(\mathbf{v}^*,\boldsymbol{\Delta\beta}^*,\boldsymbol{\theta}^*,\mu^*)}{\partial z} \end{cases} \tag{4.15}$$

for $t_0 \le t \le t_f$ and $i = 1,2,\dots,N$. Direct computing (4.15) yields to the equations (4.13). According to the optimality condition $H(\mathbf{v}^*,\boldsymbol{\Delta\beta}^*,\boldsymbol{\theta}^*,\mu^*) = \min_{\boldsymbol{\Delta\beta}_i} H(\mathbf{v}^*,\boldsymbol{\Delta\beta},\boldsymbol{\theta}^*,\mu^*)$, we obtain that the optimal control $\Delta\beta_i^*(t)$ for $t_0 \le t \le t_f$ and $i = 1,2,\dots,N$ obeys

$$\frac{\partial H(\mathbf{v}^*,\boldsymbol{\Delta\beta},\boldsymbol{\theta}^*,\mu^*)}{\partial u_i} = \mu^* \frac{dg(\Delta\beta_i^*)}{d\Delta\beta_i^*} + v_i^* \sum_{j=1}^{N} a_{ij}(1-v_j^*)\theta_j^* = 0 \tag{4.16}$$

The optimality conditions include the state equations (4.9), the costate equations (4.13), and the stationary equations (4.14). ☐

The method of Adapted Forward Backward Sweep [89] with a Runge-Kutta fourth order scheme is applied to solve the optimality system. The convergence of this method is given in [91].

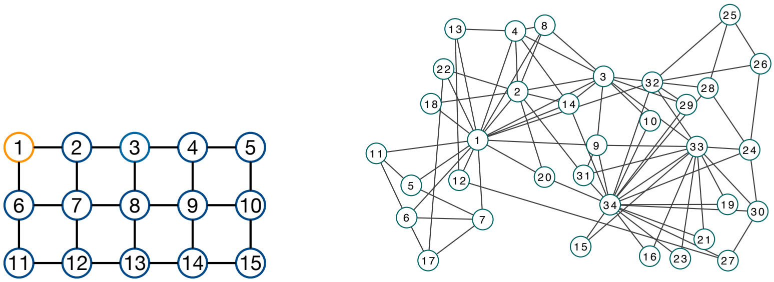
4.5. NUMERICAL RESULTS AND DISCUSSION

In this section, we investigate the behaviors of the induced SIS spreading by numerical results in some artificial and real networks. We define the quadratic cost function as $g(\Delta\beta_i) = (\Delta\beta_i)^2$ in the static method and $g(\Delta\beta_i(t)) = (\Delta\beta_i(t))^2$ in the dynamic method for $i = 1,2,\dots,N$. Since the control inputs are generally related to the external force or the electric current, the cumulative quadratic cost can be interpreted as the control energy [92] [93]. Applying the quadratic cost function also helps to compare the behavior of the induced spreading model with other existent models, e.g., LQR model. In order to guarantee the induced SIS spreading without extinction, we set the original constant infection rate $\hat{\beta} = \tau_c^{(1)} = \frac{1}{\lambda_1}$, and the infection probability $v_i(t) \approx 0$ for all nodes. Thus, the additional cost C for the induced spreading leads to a positive payoff on the infection probabilities of nodes.

4.5.1. NUMERICAL RESULTS IN THE STATIC OPTIMIZATION

PAYOFF VERSUS COST BUDGET FOR A SINGLE TARGET NODE

We first investigate the impact of the cost budget C for targeting a single node in the static model. Fig. 4.2 presents the payoff $B = J(\boldsymbol{\Delta\beta}^*)$ as a function of the cost budget C for a single target node in the lattice network $L_{3\times5}$ in Fig. 4.1a. Fig. 4.2 shows that the payoff B of the target node for the same cost budget C depends on the topological properties of the target node. The target node with a larger degree (e.g., node 2 and 3) can obtain a higher payoff for the same cost budget C.



(a) $L_{3\times5}$ (b) The Karate network

Figure 4.1: Illustration of the network topologies. (a) The lattice network $L_{3\times5}$ with $N = 15$ nodes. (b) The Karate network with $N = 34$ nodes.

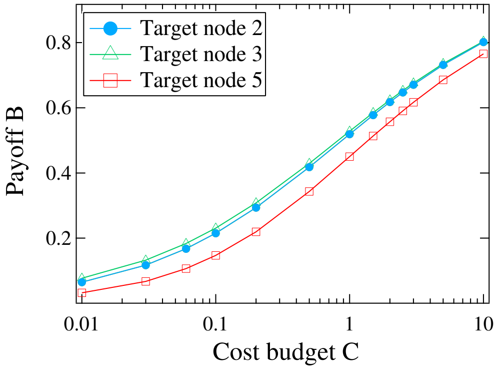


Figure 4.2: The payoff $B = J(\boldsymbol{\Delta\beta}^*)$ in the static optimization as a function of the cost budget C for a single target node in the lattice network $L_{3\times5}$.

COST ALLOCATION IN STATIC OPTIMIZATION

We apply our methods to the Karate network [94] with $N = 34$ nodes, as illustrated in Fig. 4.1b. Figure 4.3 compares the time-dependent payoff $B(t)$ after the optimal cost allocation in the exact Markovian model and NIMFA. The payoffs $B(t)$ in both models follow a similar behavior, and the gap in the payoff between NIMFA and the exact model decreases for a larger cost budget C. Since NIMFA usually provides an upper bound of infection probability in the SIS spreading process [26], the exact payoff is also upper-bounded by the payoff in NIMFA.

For the static optimization, it is of practical significance to investigate the cost allocation on the nodes in the network. We define the average target distance $\bar{h}_i$ as the mean of all the minimum hops h_{ij} from node i to the target nodes $j \in \mathcal{S}$, i.e. $\bar{h}_i = \frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}} h_{ij}$, where $|\mathcal{S}|$ denotes the number of targets. Since the infection rate β_j does not directly influence the infection probability $v_{j\infty}$ of node j, the actual minimum hops of the influence on infection probability $v_{j\infty}$ by the infection rate β_j is 2 hops. Hence, we set $h_{jj} = 2$ instead of $h_{jj} = 0$ to compute the average target distance $\bar{h}_j$ for

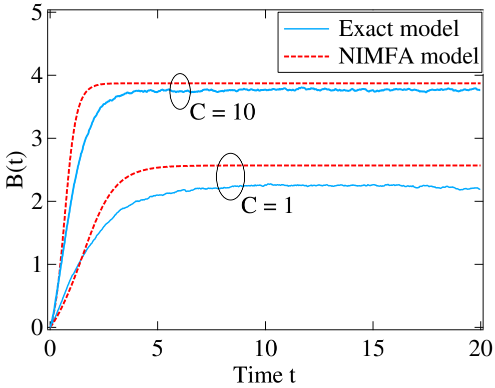


Figure 4.3: The time-dependent payoff $B(t)$ after the static optimization with a determined cost budget C for $|\mathcal{S}| = 5$ randomly selected target nodes in the Karate network. The spreading processes start from the initial spreader 34. The payoff $B(t)$ in the exact model is obtained from the simulations based on Gillespie algorithm [35] by averaging 10^4 realizations.

the target node $j \in \mathcal{S}$.

Fig. 4.4 shows the relation between the optimal rate increment $\Delta\beta_i^*$ and the topological properties of node i, e.g., the degree d_i and the average distance $\bar{h}_i$, for different fraction of targets $|\mathcal{S}|$. We illustrate that both the degree d_i and the average distance $\bar{h}_i$ are highly related to the cost allocation on nodes. Specifically, the node with a relatively high degree d_i and a shorter average distance $\bar{h}_i$ to the target nodes usually has a larger infection rate increment $\Delta\beta_i^*$. The steady-state infection probability [26] approximates

$$v_{i\infty} \approx 1 - \frac{1}{1 + \sum_{j=1}^{N} a_{ij}\beta_j \left(1 - \frac{1}{1+\sum_{k=1}^{N} a_{jk}\beta_k}\right)} \approx 1 - \frac{1}{1 + \sum_{j=1}^{N} a_{ij}\beta_j \left(1 - \frac{1}{1+d_j\beta_k}\right)} \tag{4.17}$$

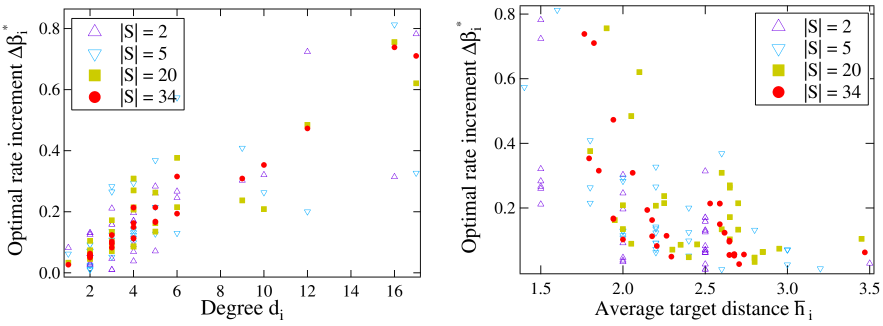
assuming that the infection rate β_k of the neighbors of node j are the same. The infection rate increment on the node with a larger degree d_j among the neighbors of node i could provide a larger payoff on the infection probability $v_{i\infty}$ of the target node i. If the fraction of target nodes $\frac{|\mathcal{S}|}{N}$ is relatively large, the cost allocated on the node with a larger degree can benefit more neighbors of this node, which leads to a stronger correlation between the infection rate increment $\Delta\beta_i^*$ and the degree d_i.

Figure 4.5 shows the relation between the optimal rate increment $\Delta\beta_i^*$ on node i and the degree d_i in Les Misérables network [71] and the dolphins network [95]. Figure 4.5 shows that the correlation between the optimal rate increment $\Delta\beta_i^*$ and the degree d_i usually becomes stronger with the increasing number of target nodes $|\mathcal{S}|$.

4.5.2. NUMERICAL RESULTS IN THE DYNAMIC OPTIMIZATION

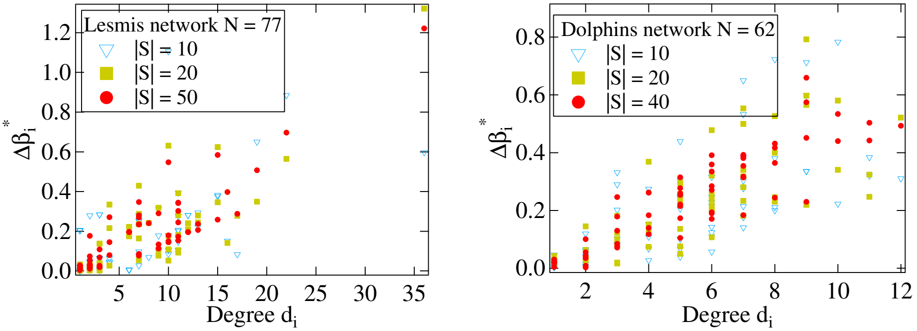
INDUCED STRATEGY FOR A SINGLE TARGET NODE

Dynamic optimization is concerned with the spreading trajectory for steering the viruses from the initial spreader to the target nodes. We first investigate the behavior of the



(a) Degree (b) Average target distance

Figure 4.4: (a) The relation between the optimal rate increment $\Delta\beta_i^*$ on node i and the degree d_i for different number of target nodes $|\mathcal{S}|$. (b) The relation between the optimal rate increment $\Delta\beta_i^*$ on node i and the average target distance $\bar{h}_i$. The target nodes are randomly selected in the Karate network with $N = 34$ nodes.

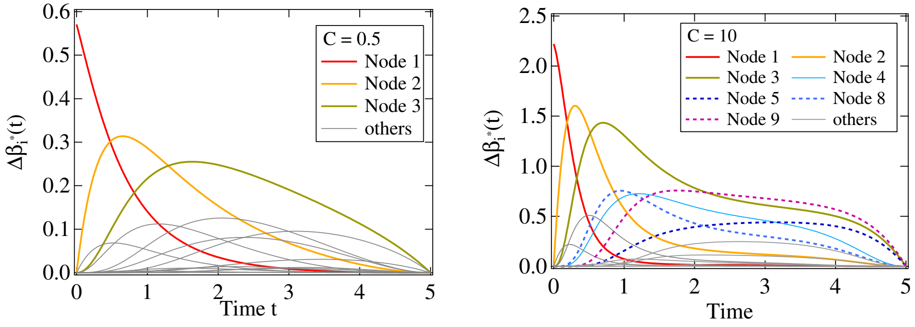


(a) Les Misérables network (b) Dolphins network

Figure 4.5: The relation between the optimal rate increment $\Delta\beta_i^*$ on node i and the degree d_i for different number of target nodes $|\mathcal{S}|$. The target nodes are randomly selected.

optimal induced spreading for a single target node. Fig. 4.6 shows the optimal control $\Delta\beta^*(t)$ in the lattice network $L_{3\times5}$ in Fig. 4.1a with the single target node 4 and the initial spreader 1, which illustrates that the behavior of the optimal rate increment $\boldsymbol{\Delta\beta}^*(t)$ depends on the cost budget C. For a small cost C (e.g., $C = 0.5$ in Fig. 4.6a), most of the cost budget is allocated to the nodes on the shortest paths from the initial spreader to the target node. For a large cost C (e.g., $C = 10$ in Fig. 4.6b), the time-dependent optimal control $\boldsymbol{\Delta\beta}^*(t)$ can be divided into two periods: first steering the viruses from the initial spreader to the target node, and then the control inputs on the neighbors (e.g. $\Delta\beta_5^*(t), \Delta\beta_8^*(t), \Delta\beta_9^*(t)$) of the target node 1 stay in a meta-steady state (e.g., $t = 1.5 - 4$) to maintain the infection probability of the target node.

We define the payoff $B = J(\boldsymbol{\Delta\beta}^*(t))$ with the optimal control $\boldsymbol{\Delta\beta}^*(t)$ for the SIS induced spreading. Fig. 4.7 shows the payoff B as a function of the cost budget C for a single target node in the lattice network $L_{3\times5}$. The logarithmic payoff $\log B$ increase faster than a linear function with the logarithmic cost budget $\log C$ for small cost budgets C, while the sub-figure in Fig. 4.7 shows the relation $B \sim \log C$ for larger cost budgets C. This result is in agreement with the above discussion on the behavior of the optimal control $\boldsymbol{\Delta\beta}^*(t)$. The small cost budget mainly contributes to steering the viruses from the initial spreader to the target nodes, and the payoff increases relatively faster. The larger cost budget is mainly allocated to the neighbors of the target nodes, and the payoff presents diminishing returns, which approximates the induced strategy in the static optimization (e.g., Fig. 4.2).



(a) C = 0.5 (b) C = 10

Figure 4.6: The optimal control $\boldsymbol{\Delta\beta}^*(t)$ for the target node 4 with the initial spreader 1 in the lattice network L_{3*5}. The initial time is $t_0 = 0$ and the final time is $t_f = 5$. (a) The optimal control $\boldsymbol{\Delta\beta}^*(t)$ for the cost budget $C = 0.5$; (b) The optimal control $\boldsymbol{\Delta\beta}^*(t)$ for the cost budget $C = 10$.

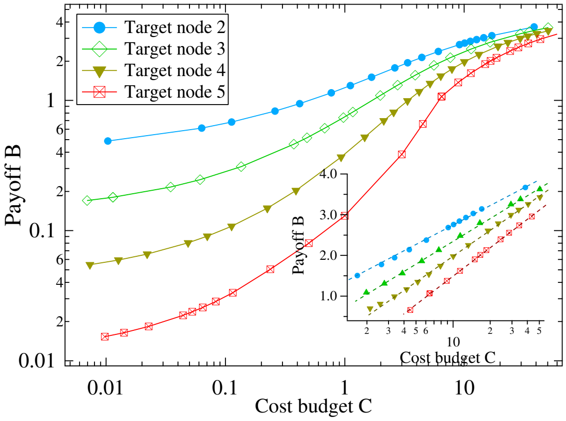
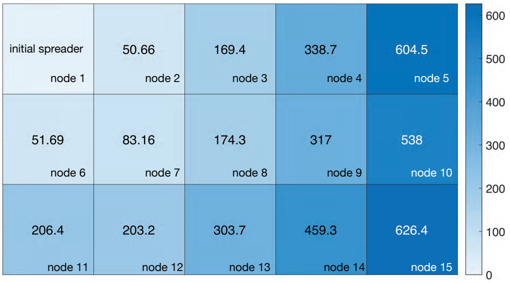
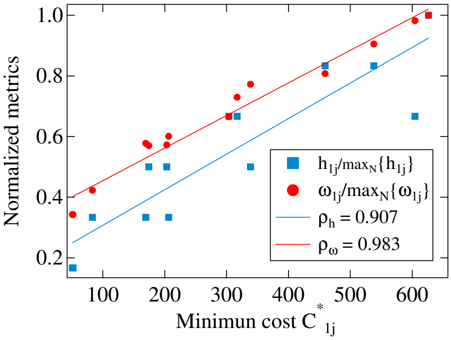


Figure 4.7: The payoff $B = J(\boldsymbol{\Delta\beta}^*(t))$ as a function of the cost C with the initial spreader 1 for a single target node in the lattice network L_{3*5}.



initial spreader node 1	50.66 node 2	169.4 node 3	338.7 node 4	604.5 node 5
51.69 node 6	83.16 node 7	174.3 node 8	317 node 9	538 node 10
206.4 node 11	203.2 node 12	303.7 node 13	459.3 node 14	626.4 node 15

(a) Minimum cost C^*_{1j}



(b) Relation between the minimum cost and the normalized metrics

Figure 4.8: (a) The minimum cost C^*_{1j} to make the target node v_j for $j = 2,3,\ldots,15$ be infected at least once in the time interval $[0,5]$, i.e., $\int_0^5 v_j(t)dt \geq 1$, in the Lattice network $L_{3\times5}$. The spreading starts from the initial spreader 1; (b) The relation between the minimum cost C^*_{1j} and the normalized shortest distance $\frac{h_{1j}}{\max_{j\in N} h_{1j}}$ as well as the normalized effective resistance $\frac{\omega_{1j}}{\max_{j\in N}\omega_{1j}}$.

COST SCALING WITH THE TOPOLOGICAL PROPERTIES OF A SINGLE TARGET NODE

Further, we investigate the cost scaling with the topological properties of a single target node. Fig. 4.8 shows the minimum cost C^*_{ij} that makes the target node v_j be infected at least once in the time interval $[0,5]$, i.e., the cumulative infection probability $J = \int_0^5 v_j(t)dt \geq 1$, in the lattice $L_{3\times5}$ with the initial infection node i. Intuitively, the minimum cost C^*_{ij} depends on the shortest distance (minimum hops) h_{ij} from the initial spreader i to the target j because all the nodes on the paths should be allocated some cost to steer the viruses to the target. Meanwhile, Fig. 4.8a illustrates that the minimum cost C^*_{ij} is also coupled to the number of paths from the initial spreader i to the target node j, i.e., a larger number of paths leads to a less cost C^*_{ij}.

Inspired by Thompson's principle [96] that the minimum energy dissipation is related to the effective resistance in electric circuits, we introduce the effective resistance

ω_{ij} between node i and node j as $\omega_{ij} = (Q^{\dagger}_{ii} + Q^{\dagger}_{jj} - 2(Q^{\dagger}_{ij}))$, where $Q^{\dagger}$ is the pseudoinverse of the Laplacian matrix Q of the network topology. Van Mieghem *et al.* [64] shows that the best conducting node j in a graph as the minimizer of the diagonal element $Q^{\dagger}_{jj}$ of the pseudoinverse matrix $Q^{\dagger}$. Fig. 4.8b shows that the relation between the optimal cost C^{*}_{1j} with the initial spreader 1 and the normalized shortest distance $\frac{h_{1j}}{\max_{j\in N} h_{1j}}$ as well as the normalized effective resistance $\frac{\omega_{1j}}{\max_{j\in N}\omega_{1j}}$. We obtain that the Pearson correlation coefficient ρ_h between the minimum cost C^{*}_{1j} and the hops h_{1j} is equal to $\rho_h = 0.907$. For the Pearson correlation coefficient ρ_ω between the minimum cost C^{*}_{1j} and the effective resistance ω_{1j}, we obtain $\rho_\omega = 0.983$. This demonstrates that the effective resistance is a good metric for the cost scaling for a single target node in the induced spreading.

COST SCALING WITH THE FRACTION OF TARGET NODES

In a linear system, the optimal control energy (cost) C^{*}_{max} for targeted controlling in the worst case [97] has the scaling equation $\log C^{*}_{max} \sim \frac{|\mathcal{S}|}{N}$, which implies that the cost C^{*}_{max} increases sharply with the increasing fraction $s = \frac{|\mathcal{S}|}{N}$ of target nodes. We will show that the behavior of energy scaling with the fraction of target nodes in the SIS process is different. We compute the optimal cost $C^{*}(s)$ subject to the constraint on the average infection frequency $\frac{1}{|\mathcal{S}|}\int_{t_0}^{t_f}\sum_{i\in\mathcal{S}} v_i(t)dt \geq 1$ of $|\mathcal{S}|$ target nodes in the SIS spreading process. Then, we can obtain the average optimal cost $E[C^{*}(s)]$ with randomly selected $|\mathcal{S}|$ target nodes in the network for multiple realizations.

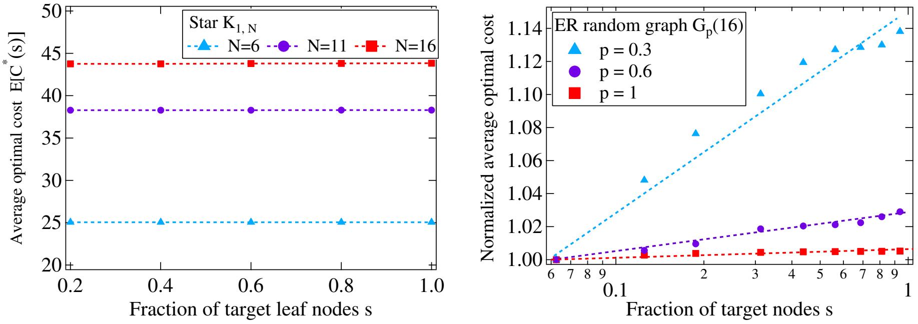
In a star network $K_{1,N}$ with an initial central spreader, the infection rate increment $\Delta\beta^{*}_{1}(t)$ on the central node for a leaf target could steer the viruses to other leaves simultaneously. Thus, the increasing fraction $s = \frac{|\mathcal{S}|}{N}$ of the target leaves influences little on the required average minimum cost $E[C^{*}(s)]$, as shown in Fig. 4.9a. Fig. 4.9b illustrates that the normalized average cost

$$\frac{E[C^{*}(s)]}{E[C^{*}(1/N)]} \sim \eta \log\frac{|\mathcal{S}|}{N} \tag{4.18}$$

in the Erdős-Rényi (ER) random network $G_p(N)$ with link density p, where η is a constant. The cost scaling law (4.18) is different from targeting control in linear control systems [97]. In linear systems, the control input for controlling a target could introduce perturbances for another target, which leads to more additional effort to control multiple targets. In the spreading process, the cost allocated on one node always benefit multiple targets by steering the viruses to them. Further, Fig. 4.9b shows that the constant η in (4.18) is coupled to the network topology and decreases with the increasing link density p in the ER random network. More nodes can benefit from the cost allocated on a single node in a denser network with a higher average degree. Thus, the less additional cost is required to steer the virus to additional targets, which translates to a smaller η.

4.5.3. COMPARISON BETWEEN THE STATIC AND THE DYNAMIC

Two optimization methods, static and dynamic optimizations, are proposed for the induced SIS spreading in the network. We now compare their performance in the Karate



(a) (b)

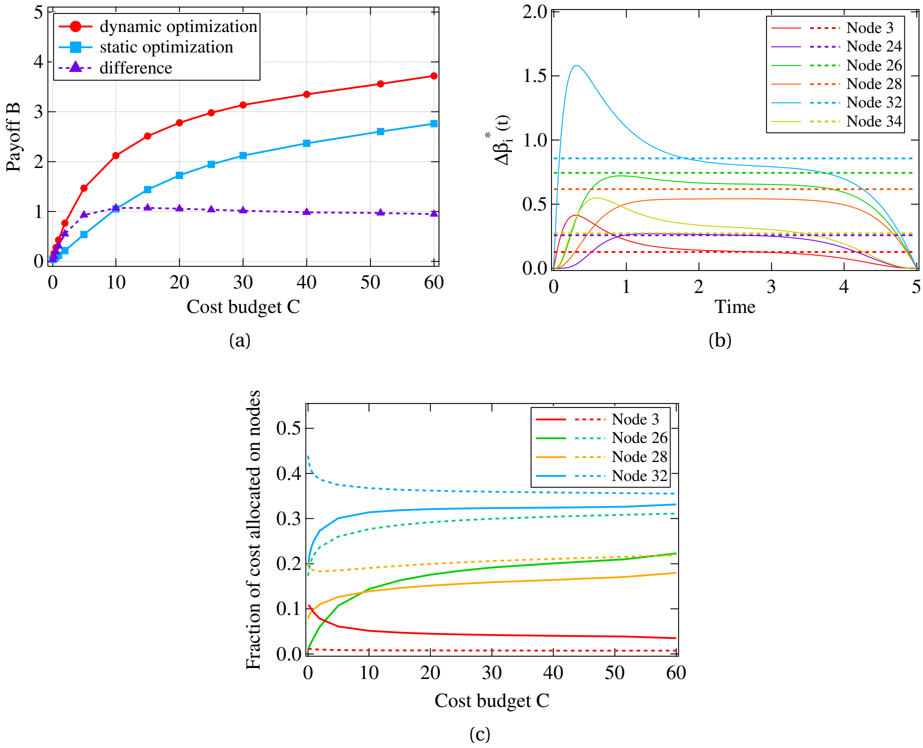
Figure 4.9: (a) The optimal cost $C^*(s)$ as a function of the fraction of leaf targets in the star $K_{1,N}$ with N leaf nodes. We set the central node as the initial spreader and the leaf nodes as the targets. (b) The normalized average cost $\frac{E[C^*(s)]}{E[C^*(1/N)]}$ as a function of the fraction of target nodes $s = \frac{|\mathcal{S}|}{N}$ in the ER random network $G_p(16)$ with the link density p. Each average optimal cost $E[C^*(s)]$ is obtained by 20 realizations. The time interval of controlling is $[t_0, t_f]$ with $t_0 = 0$ and $t_f = 5$.

network. We rewrite the constraint (4.6) in the static method as $\sum_{i=1}^{N} g(\Delta\beta_i) \leq \frac{C}{t_f - t_0}$ where C is also the cost budget in the dynamic method, and both methods have a same cost budget C in the time interval $[t_0, t_f]$. Then, we compute the payoff in the static method by $B = \int_{t_0}^{t_f} \sum_{i\in\mathcal{S}} v_i(t)$ by the NIMFA equation (4.2) with the constant optimal solution $\boldsymbol{\Delta\beta}^*(t) = \boldsymbol{\Delta\beta}^*$ for any time in the static optimization.

Fig. 4.10a shows that the dynamic method generally outperforms the static method for different cost budgets C. Specifically, the difference between both methods exhibits a maximum around $C = 10$, and then decays slowly with the cost budget C. Fig. 4.10b shows the optimal infection rate increment $\boldsymbol{\Delta\beta}^*(t)$ for the target node 25 with the initial spreader 2, where the optimal control $\boldsymbol{\Delta\beta}^*$ is time-dependent for the dynamic optimization and constant for the static optimization. The optimal control vector $\boldsymbol{\Delta\beta}^*(t)$ for the dynamic optimization (solid line) exhibits a metastable state from $t = 1-4$, and the optimal control $\boldsymbol{\Delta\beta}^*$ for the static optimization (dash line) approximates the dynamic control $\boldsymbol{\Delta\beta}^*(t)$ in the metastable state. Moreover, Fig. 4.10c shows the normalized cost allocated on the node i.e., $\frac{t_f - t_0}{C} g(\Delta\beta_i^*)$ for the static optimization and $\frac{1}{C}\int_{t_0}^{t_f} g(\Delta\beta_i^*(t))$ for the dynamic, versus the cost budget C. The difference between the normalized cost in both optimizations also becomes smaller for a larger cost budget C, which demonstrates that the dynamic induced strategy approaches the static strategy with increasing cost budget C. Thus, we suggest to apply the dynamic optimization for a limited cost budget while the performance of the static optimization is already good enough for an adequate cost budget.

4.6. RELATED WORK

Virus spread in networks has been deeply studied in recent years [12]. The previous works on the Susceptible-Infected-Susceptible (SIS) model involve the epidemic



(a)

(b)

(c)

Figure 4.10: (a) The comparison of the payoff B in the static optimization and the dynamic optimization with the cost budget C in the Karate network. The initial spreader is node 2 and the target node is node 25. The time interval of controlling is $[0,5]$. (b) The optimal control $\boldsymbol{\Delta\beta}^*$ in the static method (dash lines) and $\boldsymbol{\Delta\beta}^*(t)$ in the dynamic method (solid lines) with the cost budget $C = 10$ in the Karate network. (c) The normalized cost allocated on the node i.e., $\frac{t_f - t_0}{C} g(\Delta\beta_i^*)$ for the static optimization and $\frac{1}{C}\int_{t_0}^{t_f} g(\Delta\beta_i^*(t))$ as a function of the cost budget C.

threshold [39][40][84], the average fraction of infection nodes over time [43] and time-dependent properties in SIS processes [35][42]. The N-intertwined mean-field approximation [26] is a reasonably accurate approximation of the exact SIS epidemics on a network [27].

Some previous research has investigated epidemics control [78], which aims to stop the spreading as soon as possible. The static optimization of epidemics control can be converted to a spectral control problem, where a fixed number of resources must be optimally allocated to best mitigate the effects of a disease. Preciado *et al.* [29] propose the budget-constrained allocation problem and present a solution framework based on Geometric Programming. Some greedy strategies of link removal based on the topological properties of links for spectral control are proposed [28][98]. However, the solution by the greedy strategies may deviate much from the optimum in some worst cases [99]. The dynamic optimization, i.e., the optimal control, of a deterministic epidemic is proposed in [100][101] and solved by using Pontryagin's maximum principle. The epidemic processes on a network allow each individual to have its own state, which makes the control strategy depend on the network topology [102][103].

The problem of influence maximization [16] has a similar goal with the optimal induced spreading but different control variables. Influence maximization aims to find the optimal initial state (initial spreader) of a spreading, while the optimal induce spreading aims to adjust the parameters of the SIS model under the assumption that the initial spreaders are fixed.

4.7. CHAPTER SUMMARY

In this chapter, we explore the induced SIS spreading on networks, which aims to steer the viruses to the target nodes as much as possible by adjusting the nodal infection rates under a limited cost budget. We provide two frameworks for the optimal induced spreading: the static optimization and the dynamic optimization. We propose the algorithms for both the optimization problems and further investigate the behavior of the induced strategy by numerical results on networks with the constraint of a quadratic cost function.

In the static optimization, the optimal infection rate increment of each node is highly related to the degree and its average hops to the targets, while the degree dominates the nodal infection rate increment for a large fraction of targets. In the dynamic optimization, we show that the time-dependent optimal infection rate increment exhibits two periods for a large cost budget: steering the viruses from the initial spreader to the target, and maintaining the infection probability of the target by its neighbors. For a single target node, we show that the effective resistance could be a good metric to indicate the cost scaling. Further, we illustrate that the cost scaling with the fraction of targets has different behaviors to that of the targeted controlling in linear systems, because the cost for increasing the infection rate of one node usually benefits the infection probabilities of multiple targets. Finally, we show that the dynamic induced strategy approximates the static for a large cost budget.

Some problems of practical significance merit further study. First, the induced SIS spreading problem can be generalized to guide the infection to some target nodes while avoiding some other specific nodes as much as possible. We suspect that some results presented in this chapter, such as the cost scaling with the fraction of the targets, could change for the generalized problem. Second, the topological properties of the most efficient spreader [55] with a minimum total cost for the induced spreading is worthy of study.

5

PREVALENCE EXPANSION IN NIMFA

The N-Intertwined Mean Field Approximation (NIMFA) is a reasonably accurate approximation of the exact SIS epidemic process on a network. The average fraction of infected nodes in the NIMFA steady state, also called the steady-state prevalence, in terms of the effective infection rate can be expanded into a power series around the NIMFA epidemic threshold. In this chapter, we investigate the convergence of the steady-state prevalence Taylor expansion. We determine the radius of convergence in some special types of graphs. We also show that the radius of convergence of the steady-state prevalence expansion depends upon the network topology, in particular, the average degree of the network and the spectral gap of the adjacency matrix play a role.

This chapter is based on the published paper [104].

5.1. INTRODUCTION

EPIDEMIC models can describe virus spreading and information propagation in human activities [12]. The Susceptible-Infected-Susceptible (SIS) model is an epidemic model where each infected item can be cured, and becomes susceptible again after recovering from the infection state [38]. In the SIS epidemics on a network, the ratio between the infection rate β and the curing rate δ is called the effective infection rate $\tau = \beta/\delta$. The SIS model features a phase transition [39] around the epidemic threshold τ_c. The spreading process can reach the metastable state [35] if the effective infection rate τ is above the epidemic threshold τ_c. In the metastable state, the viruses can infect a sizeable portion of the population on average and stay in the network for a long time [45]. A first-order mean-field approximation [26] of the epidemic threshold $\tau_c^{(1)} = 1/\lambda_1$, where λ_1 is the spectral radius of the adjacency matrix A, was shown [40] to be a lower bound for the exact epidemic threshold τ_c.

The exact Markovian SIS model [26] in the network G with N nodes consists of 2^N states, which is intractable to solve for large networks. The N-Intertwined Mean-Field Approximation [105] was proposed by introducing the network topology into the deterministic model [106][107], which can approximate the exact SIS epidemics well in some networks. NIMFA approximates the exact Markovian 2^N linear equations into N non-linear differential equations under the assumption that the states of the nodes are uncorrelated. The steady-state infection probabilities $v_{i\infty}(\tau)$ of each node i in NIMFA can be expanded [108] in a power series in terms of the effective infection rate τ at NIMFA epidemic threshold $\tau_c^{(1)}$ and explicitly repeated in the Appendix C.1. Mathematically, the radius of convergence of a power series corresponds to the radius of the largest disk in which the series converges [109], which is of practical significance to validate a Taylor expansion. Practically, we can faster compute the nodal infection probability of the NIMFA steady-state by the truncated expansion with enough terms and an effective infection rate τ within the radius of convergence, instead of numerically solving the governing equation (5.1). Thus, the radius of convergence of the series, that purely depends upon the underlying topology, determines the largest effective infection rate for the Taylor expansion. However, the convergence of the Taylor expansion (C.1) in Appendix C.1 of the nodal infection probability in NIMFA is still an open question and has not been studied yet.

In this chapter, we focus on the NIMFA steady-state prevalence, which is defined as the average fraction of infected nodes in the NIMFA steady state. We numerically investigate the radius of convergence of the steady-state prevalence expansion and illustrate that the convergence of the prevalence expansion highly depends on the underlying topology. Specifically, we investigate the behaviour of the radius of convergence in some special types of graphs, e.g., regular graph, star graph, path graph, Erdős–Rényi (ER) random graph and scale-free graph, which helps to estimate the valid range of the effective infection rate τ for the steady-state NIMFA prevalence expansion (5.4). We also identify the topological properties that influence the radius of convergence in sparse networks and clustered networks.

The outline of the chapter is as follows. In Section 5.2, we briefly review the NIMFA SIS epidemics, provide related notations and describe the method for practical estimation of the radius of convergence in Section 5.2. Section 5.3 investigates the behaviour of the radius of convergence in some special types of graphs. In Section 5.4, we investigate the topological properties that dominate the radius of convergence. Finally, we summarize the chapter in Section 5.5.

5.2. Expansion of the NIMFA steady-state prevalence

5.2.1. NIMFA prevalence

The probability $v_i(t)$ that node i is infected is given in NIMFA [105] by the following first-order nonlinear differential equation

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

where $a_{ij} \in \{0,1\}$ is the entry of the adjacency matrix A of the underlying graph. If $a_{ij} = 1$, then there is a link between node i and node j, and otherwise $a_{ij} = 0$. With the infection probability vector $\boldsymbol{v}(t) = (v_1(t), v_2(t), \ldots, v_N(t))^T$, the governing equation of NIMFA in matrix form [38] is

$$\frac{d\boldsymbol{v}(t)}{dt} = \beta A \boldsymbol{v}(t) - \operatorname{diag}\big(\boldsymbol{v}(t)\big)\Big(\beta A \boldsymbol{v}(t) + \delta u\Big) \tag{5.1}$$

where $\operatorname{diag}\big(\boldsymbol{v}(t)\big)$ is a diagonal matrix with the infection probability vector $\boldsymbol{v}(t)$ and u is the all-one vector. By solving the governing equation (5.1) with the initial state $\boldsymbol{v}(0) = (v_1(0), v_2(0), \ldots, v_N(0))^T$, we can obtain the NIMFA prevalence, i.e. the average fraction of infected nodes, as $y(t) = \frac{1}{N}\sum_{i=1}^{N} v_i(t)$.

We denote the eigenvalues of the adjacency matrix A by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, and denote by x_k the eigenvector, normalized by $x_k^T x_k = 1$, corresponding to eigenvalue λ_k. The infection probabilities $v_{i\infty}$ of node i in the steady state are non-zero if the effective infection rate $\tau = \frac{\beta}{\delta}$ is above the NIMFA epidemic threshold $\tau_c^{(1)} = 1/\lambda_1$. For simplicity without the loss of generality, we set the curing rate equal to $\delta = 1$, and then the infection rate equals $\beta = \tau$. In this chapter, we focus on the steady-state prevalence $y_\infty(\tau) = \frac{1}{N}\sum_{i=1}^{N} v_{i\infty}(\tau)$ in terms of the effective infection rate τ.

5.2.2. Expansion of the NIMFA steady-state prevalence

Before discussing the Taylor expansion of the NIMFA steady-state prevalence $y_\infty(\tau)$, we note that the Taylor series of the steady-state infection probability $v_{i\infty}(\tau)$ in terms of the effective infection rate τ about the epidemic threshold $\tau_c^{(1)}$ does not exist in physical space. The physical space here defines the space of all steady-state infection probability $v_{i\infty}(\tau)$ as solution of (5.1) that are possible (i.e. a probability must be non-negative). Van Mieghem [26] obtains that the steady-state infection probability $v_{i\infty}(\tau) = 0$ for any effective infection rate $\tau \leq \tau_c^{(1)}$. The derivative of the steady-state infection probability $v_{i\infty}(\tau)$ with respect to the effective infection rate τ on the left-side of $\tau_c^{(1)}$ follows $\frac{dv_{i\infty}(\tau)}{d\tau}\Big|_{\tau=(\tau_c^{(1)})^-} = 0$. Invoking the fact [108] that $\frac{dv_i(t,\tau)}{d\tau}\Big|_{\tau=(\tau_c^{(1)})^+} = \frac{\lambda_1}{N}\frac{(x_1)_i}{\sum_{i=1}^{N}(x_1)_i^3} > 0$ for the effective infection rate $\tau > \tau_c^{(1)}$, left and right derivatives are not equal, implying that the infection probability $v_{i\infty}(\tau)$ is not an analytic function of the effective infection rate τ at epidemic threshold $\tau_c^{(1)}$. Consequently, the Taylor series does not exist. In other words, the radius of convergence R of the expansion of the NIMFA steady-state prevalence equals $R = 0$.

The expansion of the steady-state infection probability with the effective infection rate $\tau \downarrow \tau_c^{(1)}$ can be derived from

$$\beta A \boldsymbol{v}_\infty(t) - \operatorname{diag}\big(\boldsymbol{v}_\infty(t)\big)\Big(\beta A \boldsymbol{v}_\infty(t) + \delta \boldsymbol{u}\Big) = 0 \tag{5.2}$$

for the effective infection rate $\tau \to \tau_c^{(1)}$. Mathematically, we extend the solution of (5.2) to complex τ. Hence, the steady-state infection probability $v_{i\infty}(\tau)$ is allowed to be negative for $\tau < \tau_c^{(1)}$, as illustrated in Figure 5.1. Thus, the infection probability $v_{i\infty}(\tau)$ in the extended mathematical space is analytic at $s = \tau^{-1} = (\tau_c^{(1)})^{-1} = \lambda_1$, where the mathematical space includes all solutions of Equation (5.2) for $v_{i\infty}(\tau)$, regarded as a complex function of τ, ignoring its probabilistic and physical meaning. Although the negative steady-state infection probability has no physical significance, the Taylor series of the steady-state infection probability in $z = (\tau_c^{(1)})^{-1} - \tau^{-1}$ around $z = 0$ still coincides with the NIMFA positive infection probability for the effective infection rate $\tau > \tau_c^{(1)}$, as shown in Figure 5.1, by analytic continuation [109, Chap.4].

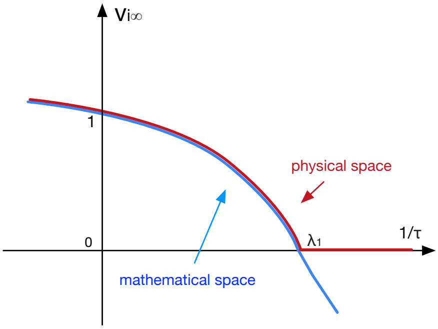


Figure 5.1: Illustration of the steady-state infection probability $v_{i\infty}$ of node i as a function of $s = \tau^{-1}$, where node i is an arbitrary node in the network. The red line represents the steady-state infection probability in the physical space, and the blue line represents the extended solution of (5.2) in the mathematical space.

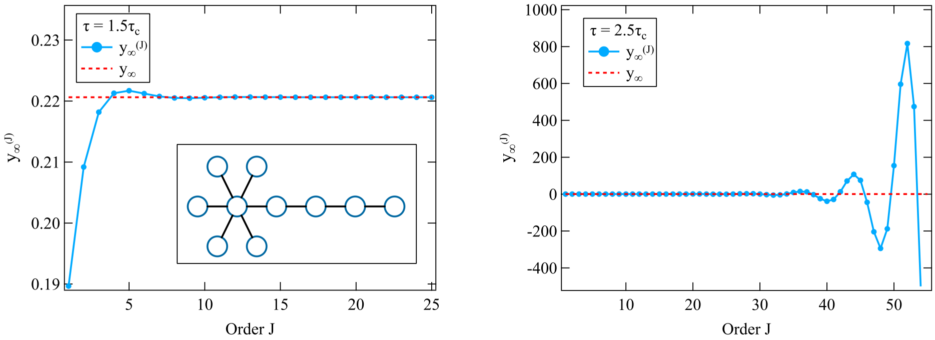
The NIMFA steady-state infection probability vector $\boldsymbol{v}_\infty(\tau)$ can be expanded in a Taylor series around $z = 0$ as

$$\boldsymbol{v}_\infty(\tau) = \sum_{j=1}^{\infty} \boldsymbol{\alpha}_j z^j = \sum_{j=1}^{\infty} \boldsymbol{\alpha}_j \big((\tau_c^{(1)})^{-1} - \tau^{-1}\big)^j. \tag{5.3}$$

where $z := (\tau_c^{(1)})^{-1} - \tau^{-1}$. The coefficient vector $\boldsymbol{\alpha}_j$ in the expansion equals

$$\boldsymbol{\alpha}_j = \sum_{k=1}^{N} c_j(k) x_k = X \boldsymbol{c}_j$$

where the matrix X contain as columns the eigenvector of the adjacency matrix A, and $c_j(k)$ satisfy [108] the recursions (C.2)-(C.4) in Appendix C.1. We ignore the trivial



(a) $y_\infty^{(J)}(\tau)$ with $\tau = 1.5\tau_c^{(1)}$

(b) $y_\infty^{(J)}(\tau)$ with $\tau = 2.5\tau_c^{(1)}$

Figure 5.2: Steady-state prevalence expansion $y_\infty^{(J)}(\tau)$ as a function of order J. The example graph topology is illustrated in the subplot (a). The red dash lines represent the NIMFA steady-state prevalence $y_\infty(\tau)$ following the NIMFA equation (5.2).

solution $v_{i\infty}(\tau) = 0$ with all-zero coefficients $c_j(k)$ and focus on the non-zero infection probability. The expansion of the steady-state infection probability $v_{i\infty}(\tau)$ with the recurrence relation of the coefficients $c_j(k)$ are derived in [108] and revisited (C.2)-(C.4) in Appendix C.1. Thus, the expansion of the NIMFA steady-state prevalence can be represented by

$$y_\infty(\tau) = \frac{1}{N} u^T \boldsymbol{v}_\infty(\tau) = \sum_{j=1}^{\infty} b_j z^j \tag{5.4}$$

where the coefficient $b_j = \frac{1}{N} u^T \boldsymbol{\alpha}_j$, and u is the all-one vector.

We define the truncated Taylor series with J terms as

$$y_\infty^{(J)}(\tau) = \sum_{j=1}^{J} b_j z^j \tag{5.5}$$

and investigate the accuracy and convergence of the expansion $y_\infty^{(J)}(\tau)$. The steady-state prevalence expansion $y_\infty(\tau)$ converges, if for every arbitrarily small number $\epsilon > 0$, there exists an order J_c such that $|y_\infty^{(J)} - y_\infty| < \epsilon$ for all $J > J_c$. The radius of convergence R is a nonnegative real number or ∞ such that the series (5.4) converges if $|z| = |(\tau_c^{(1)})^{-1} - \tau^{-1}| < R$ and diverges if $|z| = |(\tau_c^{(1)})^{-1} - \tau^{-1}| > R$ in the complex z-plane. Figure 5.2 shows the NIMFA steady-state prevalence expansion $y_\infty^{(J)}(\tau)$ as a function of J for two different effective infection rate τ in an example graph. Figure 5.2 illustrates that the prevalence expansion $y_\infty^{(J)}(\tau)$ converges to the NIMFA steady-state prevalence $y_\infty(\tau)$ if the effective infection rate τ is small (e.g., $\tau < 1.5\tau_c^{(1)}$), but diverges if τ is large enough (e.g., $\tau > 2.5\tau_c^{(1)}$) along with a large term z.

5.2.3. ESTIMATION OF THE RADIUS OF CONVERGENCE

For the power series $y_\infty(\tau) = \sum_{j=1}^{\infty} b_j z^j$ where $z = (\tau_c^{(1)})^{-1} - \tau^{-1}$ expanded at $z = 0$, the radius of convergence R in the complex z-plane is given by [109]

$$R^{-1} = \lim_{j\to\infty} \left| \frac{b_{j+1}}{b_j} \right| = \limsup_{j\to\infty} \sqrt[j]{|b_j|} \tag{5.6}$$

Unfortunately, the radius of convergence R of the prevalence expansion in a general graph is intractable to derive analytically. The method of Domb-Sykes plot and Mercer-Roberts plots [110], which are usually applied to estimate the radius of convergence of a series if the sign of the coefficients b_j follows a pattern, is also impractical since the coefficients b_j could present an unpredictable sign pattern in a general graph.

In the following sections, we assess the convergence of the expansion numerically. We regard that the prevalence expansion (5.4) is convergent if the prevalence $y_\infty^{(J)}$ with a large enough order J approximates the NIMFA steady-state prevalence y_∞^* solved by the equations (5.1), i.e., $|y_\infty^{(J)} - y_\infty| < \epsilon$ for all $J > J_c$. In the following simulations, we set $J_c = 150$ and $\epsilon = 0.001 y_\infty$. We have numerically verified that the setting of $J_c = 150$ is large enough to assess the convergence of the expansions in the following simulations.

5.3. RADIUS OF CONVERGENCE R OF THE EXPANSION IN SOME SPECIFIC GRAPHS

We first investigate the behaviour of the coefficients b_j and the radius of convergence R of the expansion of the steady-state prevalence in some special types of graphs, e.g., complete graphs, regular graphs, star graphs, path graphs, Erdős–Rényi random graphs and scale-free graphs.

5.3.1. COMPLETE GRAPHS AND REGULAR GRAPHS

In the complete graph K_N, the NIMFA steady-state prevalence is

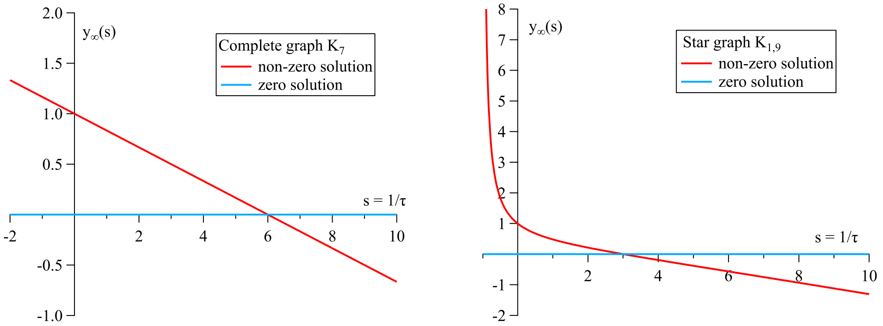
$$y_\infty(s) = 1 - \frac{1}{\tau(N-1)} = \frac{1}{\lambda_1}\left((\tau_c^{(1)})^{-1} - s\right) \tag{5.7}$$

where $s = \tau^{-1}$. The prevalence (5.7) is analytic in the whole complex plane. Hence, the radius of convergence R is infinity. The coefficient $b_1 = \frac{\sum_{i=1}^{N} (x_1)_i}{N} c_1(1) = (\lambda_1 \sum_{i=1}^{N} (x_1)_i^3)^{-1} = \frac{1}{N-1}$, and $b_j = 0$ for $j > 1$, which is consistent with (5.7). Figure 5.3a illustrates that the exact solutions of the steady-state governing equation (5.2) in a complete graph has two branches (illustrated in Figure 5.3a), and the branch corresponding to the non-trivial real solution (red line) is analytic for the reciprocal of any effective infection rate $s = \tau^{-1}$.

In the r-regular graph with the average degree $E[D] = r$, the steady-state prevalence is $y_\infty(s) = \frac{1}{r}((\tau_c^{(1)})^{-1} - s)$, which implies the steady-state prevalence expansion (5.4) also has an infinite radius of convergence.

5.3.2. STAR GRAPHS $K_{1,N-1}$

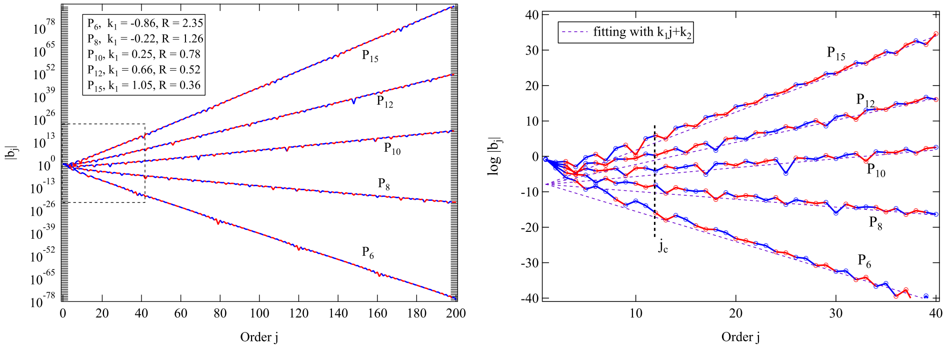
In the star graph $K_{1,N-1}$ with $N-1$ leaf, the NIMFA steady-state infection probability [49] for the center node equals $v_{c\infty} = \frac{(N-1)-s^2}{(N-1)+s}$ and $v_{c\infty} = 0$, while the NIMFA infection



(a) y^*_∞ in K_7 (b) y^*_∞ in $K_{1,9}$

Figure 5.3: Exact steady-state prevalence y^*_∞ by the solution of the equation (5.2) in the complete graph K_7 and the star graph $K_{1,9}$.

probability for the leaf node equal $v_{l\infty} = \frac{(N-1)-s^2}{(N-1)(s+1)}$ and $v_{l\infty} = 0$, where $s = \tau^{-1}$. Thus, the steady-state prevalence follows $y_\infty = \frac{1}{N}\left(\frac{(N-1)-s^2}{(N-1)+s} + \frac{(N-1)-s^2}{(s+1)}\right)$, which has nearest singularity $s = -1$. The radius of convergence R of the expansion (5.4) equals $R = \sqrt{N-1}+1$ in a star graph, which implies the expansion (5.4) converges for any effective infection rate $\tau > \tau_c^{(1)}$. Figure 5.3b illustrates the NIMFA prevalence $y_\infty(\tau)$ in a star graph, where the non-zero solution corresponds to the expansion (5.2) has the singularity at $s = -1$.



(a) (b)

Figure 5.4: Coefficient $|b_j|$ as a function of order j in path graphs P_N with N nodes. The red markers represent positive coefficients, and the blue markers represent negative. The function $|b_j|$ is fitted by the function $e^{k_1 j + k_2}$. Subplot (b) shows the function $\log|b_j|$ for small order j.

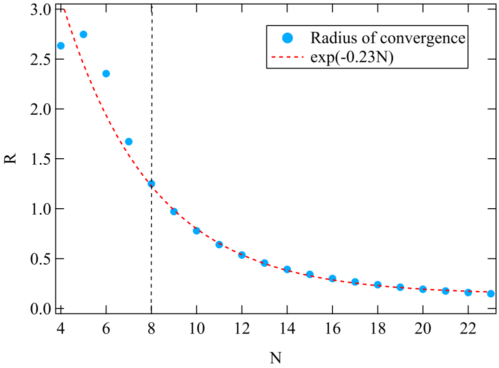


Figure 5.5: Radius of convergence R of the steady-state prevalence expansion as a function of network size N in path graphs P_N.

5.3.3. PATH GRAPHS P_N

Since the computation of the NIMFA prevalence in a path graph is intractable, we numerically investigate the radius of convergence of the prevalence expansion. Figure 5.4 shows the function of the absolute coefficients $|b_j|$ in the prevalence expansion in path graphs with different network size N, while the period and the sign pattern of the coefficients b_j is too complicated to predict. Figure 5.4b shows that the absolute coefficients $|b_j|$ resembles an exponential function $|b_j| \approx e^{k_1 j + k_2}$ if the order j is larger than a critical order j_c. Hence, the radius of convergence R can be estimated by $R \approx e^{-k_1}$, which is almost the same as the radius of convergence estimated in Section 5.2.3. Further, the NIMFA prevalence with $z = (\tau_c^{(1)})^{-1} - \tau^{-1} < e^{-k_1}$ can be estimated by

$$y_\infty^*(\tau) \approx \sum_{j=1}^{j_c-1} b_j z^j + \sum_{j=j_c}^{\infty} e^{k_1 j + k_2} z^j = \sum_{j=1}^{j_c-1} b_j z^j + e^{k_2} \frac{(e^{k_1} z)^{j_c}}{1 - e^{k_1} z} \tag{5.8}$$

Figure 5.5 shows that the radius of convergence R decreases with increasing network size N in path graphs P_N but not monotonically, i.e., R reaches the maxima at $N = 5$. The radius of convergence R decays as an exponential function of the network size if $N > 8$, and tends to be very small in path networks with a large size, which implies that the effective infection rate τ must be extremely close to the epidemic threshold $\tau_c^{(1)}$ to converge the prevalence expansion (5.4) in large path networks.

5.3.4. ER RANDOM GRAPHS

We first present the behaviour of the coefficient b_j in the prevalence expansion in ER random graphs $G_p(N)$ with link density p. Figure 5.6 shows that the coefficients b_j can be negative for a small link density p. The coefficients b_j are always positive if the link density p is large enough, which implies that the NIMFA prevalence function in ER graphs has a dominating positive singularity [111]. The absolute coefficients $|b_j|$ also follow an exponential function $|b_j| \sim e^{kj}$, and the radius of convergence R can be estimated by $R \approx e^{-k}$. Figure 5.7a shows the distribution of the radius of convergence R in ER random graph $G_p(N)$. The distribution type does not differ much for different link

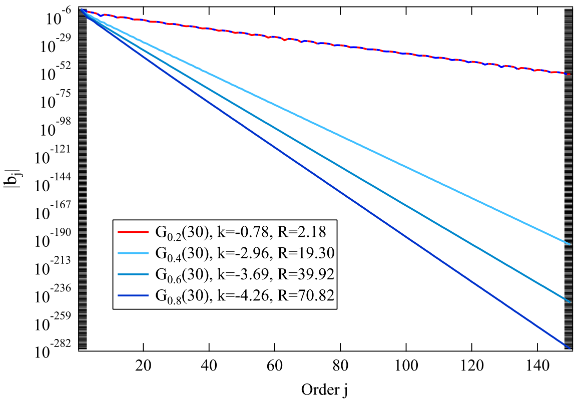
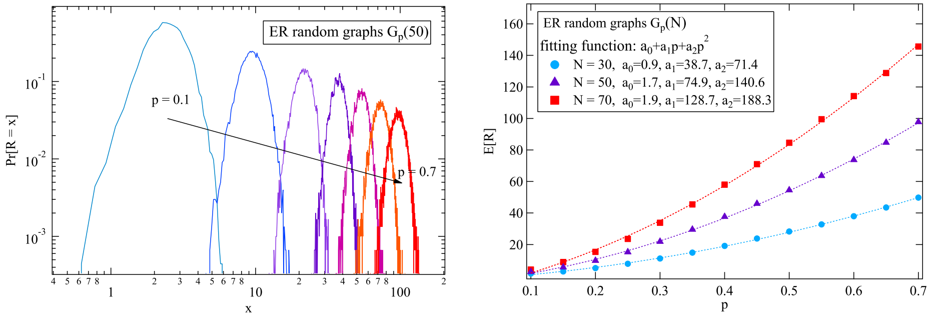


Figure 5.6: Coefficient $|b_j|$ as a function of order j in ER random graphs $G_p(N)$ with $N = 30$ nodes and link density p. The red markers represent positive coefficients and the blue markers represent negative. The slope of the function $\log|b_j|$ with order j is denoted by k.

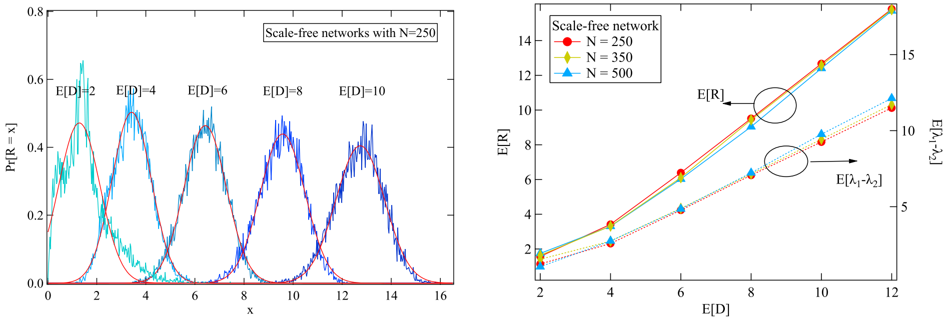


(a) (b)

Figure 5.7: Subplot (a): Histogram of the radius of convergence R in ER random graphs $G_p(30)$ based on 2×10^4 realizations for each link density p. Subplot (b): Average radius of convergence $E[D]$ in ER random graphs $G_p(N)$ as function of link density p for different network size N

density p, while the tail of the distribution becomes a little heavier with the increasing link density p. A larger link density p implies more regular ER random graphs, which leads to a higher probability for a large radius of convergence R.

Figure 5.7b shows the average radius of convergence $E[R]$ in ER random graphs increases with the link density p, and resembles a quadratic function $E[R] = a_0 + a_1 p + a_2 p^2$ for a network size N. If the link density is large enough, e.g., $p > 0.7$ in $G_p(70)$, the radius of convergence of the prevalence expansion in ER random graphs can be extremely large with a high probability because the ER graph tends to a regular graph.



(a) (b)

Figure 5.8: Subplot (a): Histogram of the radius of convergence R in scale-free graphs with network size $N = 250$ based on 2×10^4 realizations for each average degree $E[D]$. Subplot (b): Average radius of convergence $E[D]$ in scale-free graphs as a function of average degree $E[D]$ for different network size N based on 10^4 realizations. Average spectral gap $E[\lambda_1 - \lambda_2]$ in scale-free graphs as a function of average degree $E[D]$ for different network size N based on 2×10^3 realizations.

5.3.5. SCALE-FREE GRAPHS

Figure 5.8a shows the histogram of the radius of convergence in scale-free networks with network size $N = 250$ and different average degree $E[D]$. The radius of convergence in scale-free networks resembles a Gaussian distribution for the average degree $E[D] > 2$. Figure 5.8b shows the average radius of convergence $E[R]$ increases with the average degree $E[D]$, which is in line with the analysis in ER random graphs that the denser and more regular graph usually leads to a higher radius of convergence.

We observe that the network size N almost does not influence the average radius of convergence $E[R]$, which reminds invariant topological properties with the network size N in scale-free networks. Figure 5.8b shows that the average spectral gap $E[\lambda_1 - \lambda_2]$ also differs little with increasing the network size N in scale-free networks, which hints that the radius of convergence R in scale free graphs may be dominated by the spectral gap $\lambda_1 - \lambda_2$.

5.4. EFFECT OF THE TOPOLOGICAL PROPERTIES ON R

In the last section, we illustrate that the prevalence expansion in a dense or regular network usually has a large radius of convergence R. The prevalence expansion is valid for any effective infection rate τ if the radius of convergence $R > \lambda_1$. We investigate the relatively small radius convergence $R < \lambda_1$, which often occurs in sparse networks and clustered networks.

5.4.1. EFFECT OF THE NETWORK TOPOLOGY ON R

Since the radius of convergence R of the steady-state prevalence expansion (5.4) can be estimated by the inverse of the convergence order of the absolute coefficient $|b_j|$, i.e., $R = \frac{1}{\lim_{j\to\infty} \sqrt[j]{|b_j|}}$. We first investigate the behavior of the absolute coefficient $|b_j|$.

Lemma 5.1 *The convergence order of the coefficients* $|b_j| = |u^T \boldsymbol{\alpha}_j|$ *in the prevalence expansion* (5.4) *is upper bounded by the convergence order of the norm of coefficients* $||\boldsymbol{\alpha}_j||$ *and the norm of coefficients* $||\boldsymbol{c}_j||$*, i.e.*

$$\lim_{j\to\infty} |b_j|^{\frac{1}{j}} \le \lim_{j\to\infty} ||\boldsymbol{\alpha}_j||^{\frac{1}{j}} = \lim_{j\to\infty} ||\boldsymbol{c}_j||^{\frac{1}{j}} \tag{5.9}$$

where the coefficient vector $\boldsymbol{\alpha}_j = X\boldsymbol{c}_j$*.*

Proof: See Appendix C.2. ☐

Then we investigate the convergence order of the coefficient $|b_j|$ by making use of the coefficient vector $\boldsymbol{c}_j$. Defining $z = (\tau_c^{(1)})^{-1} - \tau^{-1}$ and substituting the infection probability vector $\boldsymbol{v}_\infty(\tau) = \sum_{j=1}^{\infty} \boldsymbol{\alpha}_j z^j$ into (5.2), we can obtain that

$$A\sum_{j=1}^{\infty} \boldsymbol{\alpha}_j z^j - (\lambda_1 - z)\sum_{j=1}^{\infty} \boldsymbol{\alpha}_j z^j = diag\left(\sum_{j=1}^{\infty} \boldsymbol{\alpha}_j z^j\right) A\left(\sum_{j=1}^{\infty} \boldsymbol{\alpha}_j z^j\right) \tag{5.10}$$

Equating the coefficients of the term z^j for $j \ge 2$ yields

$$\lambda_1 \boldsymbol{\alpha}_j - A\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_{j-1} - \sum_{k=1}^{j-1} diag(\boldsymbol{\alpha}_k) A\boldsymbol{\alpha}_{j-k} \tag{5.11}$$

Invoking the coefficient $\boldsymbol{\alpha}_j = X\boldsymbol{c}_j$, we multiplying X^T on both sides of (5.11) and arrive

$$(\lambda_1 I - \Lambda)\boldsymbol{c}_j = \boldsymbol{c}_{j-1} - X^T \sum_{k=1}^{j-1} diag(X\boldsymbol{c}_k) AX\boldsymbol{c}_{j-k} \tag{5.12}$$

where the adjacency matrix $A = X\Lambda X^T$, and Λ is the diagonal matrix with $\Lambda_{ii} = \lambda_i$.

For $m = 1$, we introduce the basis vector e_m and obtain that

$$(\boldsymbol{c}_j)_1 = e_1^T \left(X^T \sum_{k=1}^{j} diag(\boldsymbol{\alpha}_k) A\boldsymbol{\alpha}_{j+1-k}\right)$$

which yields that

$$\begin{aligned}|(\boldsymbol{c}_j)_1| &= \left|e_1^T \left(X^T \sum_{k=1}^{j} diag(X\boldsymbol{c}_k) AX\boldsymbol{c}_{j+1-k}\right)\right| \le \left|\left|\sum_{k=1}^{j} diag(\boldsymbol{c}_k) A\boldsymbol{c}_{j+1-k}\right|\right| \\ &\le \lambda_1 \sum_{k=1}^{j} ||(\boldsymbol{c}_k)|| \cdot ||\boldsymbol{c}_{j+1-k}|| = 2\lambda_1 ||\boldsymbol{c}_1|| \cdot ||\boldsymbol{c}_j|| + \lambda_1 \sum_{k=2}^{j-1} ||\boldsymbol{c}_k|| \cdot ||\boldsymbol{c}_{j+1-k}||\end{aligned} \tag{5.13}$$

Supposing that $\lim_{j\to\infty} |(\boldsymbol{c}_j)_1|^{\frac{1}{j}} = \max_{i\in N}\{\lim_{j\to\infty} |(\boldsymbol{c}_j)_i|^{\frac{1}{j}}\}$, i.e., $||\boldsymbol{c}_j|| \le \sqrt{N}|(\boldsymbol{c}_j)_1|$, we arrive at

$$|(\boldsymbol{c}_j)_1| \le \frac{\lambda_1}{1 - \lambda_1 ||\boldsymbol{c}_1|| \sqrt{N}} \sum_{k=2}^{j-1} ||(\boldsymbol{c}_k)|| \cdot ||\boldsymbol{c}_{j+1-k}|| \tag{5.14}$$

For $m > 1$, we similarly obtain that

$$
\begin{aligned}
|(\boldsymbol{c}_j)_m| &= \frac{1}{\lambda_1 - \lambda_m}\left| e_m^T \left(\boldsymbol{c}_{j-1} - X^T \sum_{k=1}^{j-1} diag(X\boldsymbol{c}_k)AX\boldsymbol{c}_{j-k} \right) \right| \\
&\leq \frac{1}{\lambda_1 - \lambda_2}||\boldsymbol{c}_{j-1}|| + \frac{1}{\lambda_1 - \lambda_2}\left|\left| \sum_{k=1}^{j-1} diag(\boldsymbol{c}_k)A\boldsymbol{c}_{j-k} \right|\right| \\
&\leq \frac{1}{\lambda_1 - \lambda_2}||\boldsymbol{c}_{j-1}|| + \frac{\lambda_1}{\lambda_1 - \lambda_2}\sum_{k=1}^{j-1} ||\boldsymbol{c}_k|| \cdot ||\boldsymbol{c}_{j-k}|| \qquad (5.15)
\end{aligned}
$$

The recurrence relations (5.14) and (5.15) show that the upper bound of the maximal convergence order of the coefficient $|(\boldsymbol{c}_j)_m|$, i.e. $\max_{i \in N}\{\lim_{j\to\infty}|(\boldsymbol{c}_j)_i|^{\frac{1}{j}}\}$, as well as the upper bound of $\lim_{j\to\infty}|b_j|^{\frac{1}{j}}$ and the lower bound of the radius of convergence R, are coupled to the largest eigenvalue λ_1 and the spectral gap $\lambda_1 - \lambda_2$. We further propose a heuristic lower bound of the radius of convergence of the steady-state prevalence expansion (5.4) in Appendix C.3.

From the equation of coefficients (5.11), we also obtain the inequality

$$
||\lambda_1 I - A|| \cdot ||\boldsymbol{\alpha}_j|| \geq \left|\left|(\lambda_1 I - A)\boldsymbol{\alpha}_j\right|\right| \geq \left|\left|\boldsymbol{\alpha}_{j-1} - \sum_{k=1}^{j-1} diag(\boldsymbol{\alpha}_k)A\boldsymbol{\alpha}_{j-1-k}\right|\right| \qquad (5.16)
$$

which leads to the recurrence relation

$$
||\boldsymbol{\alpha}_j|| \geq \frac{1}{\lambda_1 - \lambda_N}\left| ||\boldsymbol{\alpha}_{j-1}|| - \left|\left| \sum_{k=1}^{j-1} diag(\boldsymbol{\alpha}_k)A\boldsymbol{\alpha}_{j-1-k} \right|\right| \right| \qquad (5.17)
$$

The recurrence relation (5.17) implies that the lower bound of the norm coefficient $||\boldsymbol{\alpha}_j||$, as well as the upper bound of the radius of convergence R, is related to the maximal difference of eigenvalues $\lambda_1 - \lambda_N$.

5.4.2. NUMERICAL TESTS

Since a large number of real-world networks are sparse networks with a small average degree, we first investigate the radius of convergence R in random sparse networks with N nodes and L links, which are generated in two steps to guarantee the connectivity. First, at step 1, we generate a (uniformly chosen) random spanning tree based on a complete graph with N nodes [112]; and then, at step 2, we add additional $L - N + 1$ links, randomly and uniformly, on the complementary graph of the spanning tree. Figure 5.9 shows the correlation between the radius of convergence R and the topological properties, e.g., the largest eigenvalue λ_1, the spectral gap $\lambda_1 - \lambda_2$, the maximal difference of eigenvalues $\lambda_1 - \lambda_N$. The above three topological properties all present a high correlation with the radius of convergence, while the spectral gap $\lambda_1 - \lambda_2$ has the highest correlation coefficient.

We present the radius of convergence R in modular ER random graphs $G(N, m, p, p_m)$, where N is the number of nodes, m is the number of modules, p is the overall link density and p_m is the proportion of links within modules [113][114]. Figure 5.10 shows that

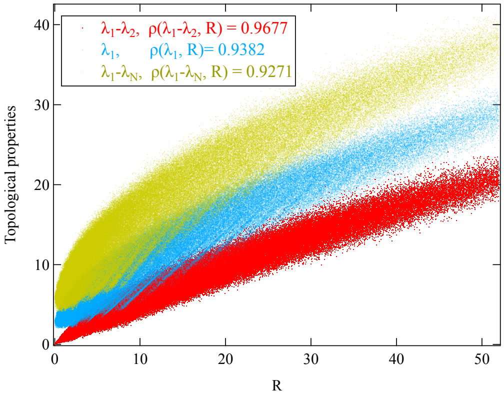


Figure 5.9: Correlation between the radius of convergence R and the topological properties, e.g., the largest eigenvalue λ_1, the spectral gap $\lambda_1 - \lambda_2$, the maximal difference of eigenvalues $\lambda_1 - \lambda_N$. The plot is based on 2×10^5 realizations of sparse random graphs. The random sparse graph is generated with the network size uniformly chosen in $N \in [10, 120]$ and the link density uniformly chosen in $p \in [\frac{2}{N}, \frac{2}{N} + 0.3]$, i.e., the additional number of links on the spanning tree is $(p - \frac{2}{N})\frac{N(N-1)}{2}$.

the correlation between the spectral gap and the radius of convergence is still high in modular ER random graphs, while the correlation between the radius of convergence R and the other properties, e.g., the largest eigenvalue λ_1 and the maximal difference of eigenvalues $\lambda_1 - \lambda_N$, degrades much.

We further explore the relation between the radius of convergence R and the topological properties in 10 real-world networks. The topologies of these networks are obtained from Newman network collection[1] and Pajek data sets[2]. Figure 5.11 shows that the spectral gap $\lambda_1 - \lambda_2$ has a stronger correlation than the largest eigenvalue λ_1 with the radius of convergence R. Moreover, the spectral gap $\lambda_1 - \lambda_2$ can act as an indicator to estimate the radius of convergence R in sparse networks.

We verify that the spectral gap of the adjacency matrix plays a critical role in the radius of convergence. The spectral gap is also an indicator of the spectral expansion[115], which describes the goodness of connectivity and Cheeger constant of a graph [116]. A larger spectral gap could increase the radius of convergence and improve the validity of the NIMFA stead-state prevalence expansion, which coincides with the fact that NIMFA approaches the exact SIS model better in well-connected networks.

5.5. CHAPTER SUMMARY

In this chapter, we investigate the convergence of the NIMFA steady-state prevalence expansion (5.4) in terms of the effective infection rate at NIMFA epidemic threshold. The network topology alters the radius of convergence R of the steady-state prevalence expansion (5.4), which is infinite in regular graphs and becomes finite in irregular graphs. The average radius of convergence increases with the density (the average

---

[1] http://www-personal.umich.edu/mejn/netdata/

[2] http://vlado.fmf.uni-lj.si/pub/networks/data/

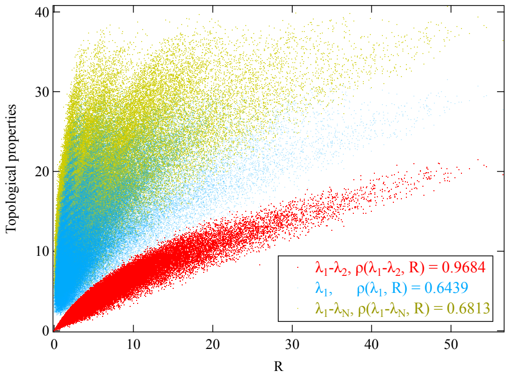
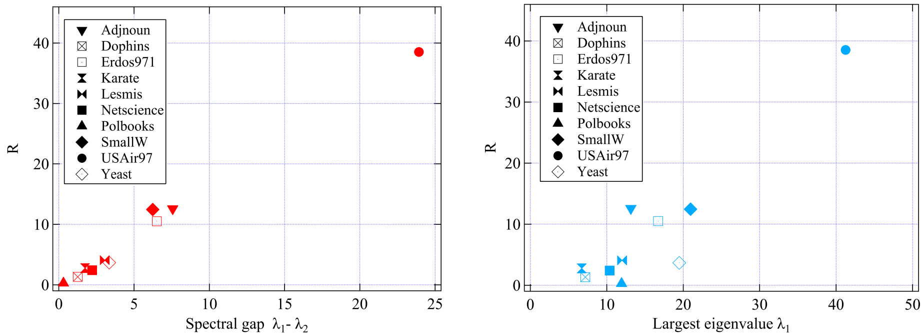


Figure 5.10: Correlation between the radius of convergence R and the topological properties, e.g., the largest eigenvalue λ_1, the spectral gap $\lambda_1 - \lambda_2$, the maximal difference of eigenvalues $\lambda_1 - \lambda_N$. The plot is based on 2×10^5 realizations of modular ER random graphs $G(N, m, p, p_m)$ with $N \in [10, 100]$, $m \in [1, 6]$, $p \in [0.05, 0.3]$ and $p_m \in [0.5, 0.9]$.



(a) Spectral gap

(b) Largest eigenvalue

Figure 5.11: The relation between the radius of convergence R and the topological properties (i.e., the spectral gap $\lambda_1 - \lambda_2$ and the largest eigenvalue λ_1) in 10 real-world networks. The networks Adjnoun, Dophins, Karate, Lesmis, Netscience, Polbooks are collected from Newman network collection, and the networks Erdos971, USAir97, and Yeast are collected from Pajek data sets. We extract the giant component of each network and regard all links as undirected and unweighted.

degree) in random graphs, e.g., ER random graph and scale-free graphs. The radius of convergence R is also coupled to the eigenvalues of the adjacency matrix, especially, a smaller spectral gap $\lambda_1 - \lambda_2$ usually decreases the radius of convergence in sparse networks and clustered networks.

Part II: Robustness of transport on networks

6

TOPOLOGICAL APPROACH TO MEASURE NETWORK RECOVERABILITY

Network recoverability refers to the ability of a network to return to a desired performance level after suffering malicious attacks or random failures. This chapter proposes a general topological approach and recoverability indicators to measure the network recoverability in two scenarios: 1) recovery of damaged connections and 2) any disconnected pair of nodes can be connected to each other. Our approach presents the effect of the random attack and recovery processes on the network performance by the robustness envelopes of realizations and the histograms of two recoverability indicators. By applying the effective graph resistance and the network efficiency as robustness metrics, we employ the proposed approach to assess 10 real-world communication networks. Numerical results verify that the network recoverability is coupled to the network topology, the robustness metric and the recovery strategy. We also show that a greedy recovery strategy could provide a near-optimal recovery performance for the investigated robustness metrics.

This chapter is based on published papers [117]

6.1. INTRODUCTION

IN communication networks, disaster-based failures and damage to optical fiber cables can partially overload data delivery, resulting in un-availability of communication services [118]. The causes for such massive failures include: human errors, malicious attacks, large-scale disasters, and environmental challenges [119]. Calculating the performance of networks under such challenges can provide significant insight into the potential damage they can incur, as well as provide a foundation for creating more robust infrastructure networks.

Network robustness is interpreted as a measure of the network's response to perturbations or challenges imposed on the network [17], which has been studied extensively in recent years. Van Mieghem *et al.* [17] propose a framework for computing topological network robustness by considering both a network topology and a service for which the network is designed. In communication networks, Cholda *et al.* [120] survey various robustness frameworks and present a general framework classification, while Pašić *et al.* [121] present the FRADIR framework that incorporates reliable network design, disaster failure modeling and protection routing. A wide range of metrics based on the underlying topology have been proposed to measure network robustness [122], and further a structural robustness comparison of several telecommunication networks under random nodal removal was presented in [123].

In a board sense, network robustness is also related to the ability of a network to return to a desired performance level after suffering malicious attacks and random failures [124]. We define such network capability as *network recoverability*[1] in this chapter. Several recovery mechanisms have been investigated under different circumstances [125], particularly in complex networks applications. Majdandzic *et al.* [21] model cascading failures and spontaneous recovery as a stochastic contiguous spreading process and exhibit a phase switching phenomenon. The recovery strategies based on the centrality metrics of network elements (e.g., nodes or links) are investigated in [124][126], which show that a centrality metric-based strategy may not exist to improve all the network performance aspects simultaneously. A progressive recovery approach [127], that consists in choosing the right sequence of links to be restored after a disaster in communication networks, proposes to maximize the weighted sum of the total flow over the entire process of recovery [128], as well as to minimize the total cost of repair under link capacity constraints [129].

Although the above studies [124]–[129] have contributed to this field, a general framework or methodology for quantifying the recovery capability of a real communication network is still lacking. In this chapter, we propose a topological approach and two recoverability indicators to measure the network recoverability in two different scenarios, link-based Scenario A and energy-based Scenario B. Specifically, Scenario A assumes that any disconnected pair of nodes can be connected to each other in the recovery process, which can describe the recovery process for logical networks. Scenario B restricts the under-repaired links to the damaged links in the attack process, which describes the recovery process for physical networks.

The proposed approach involves three concepts: the network topology, the robustness metric and the recovery strategy. For a communication network G, we apply the network efficiency E_G and the effective graph resistance r_G as the robustness metrics for case studies. The network efficiency E_G gives an indication of the efficiency of information exchange on networks under shortest path routing [130], while the effective graph resistance determines the overall diffusivity of information spreading in a communication network [64]. Besides a random recovery strategy and some strategies based on topological properties, we also consider a greedy recovery strategy. In the greedy strategy, the damaged element (a node or a link) which improves the network performance most has the highest priority to be recovered. We test our approach in 10 real telecommunication networks, including logical networks and backbone networks located in different areas, and verify that the proposed approach and the proposed recoverability indicators can assess the performance of different recovery strategies and compare the recoverability of different networks.

---

[1] Sometimes *network restoration* is used

The rest of this chapter is organized as follows: Section 6.2 introduces the topological approach for measuring the network recoverability in two scenarios. Section 6.3 presents the main concepts in the evaluation of network recoverability. The experimental results are exhibited in Section 6.4. Section 6.5 concludes the chapter.

6.2. TOPOLOGICAL APPROACH FOR MEASURING NETWORK RECOVERABILITY

In this section, we introduce an approach for measuring the network recoverability in two scenarios.

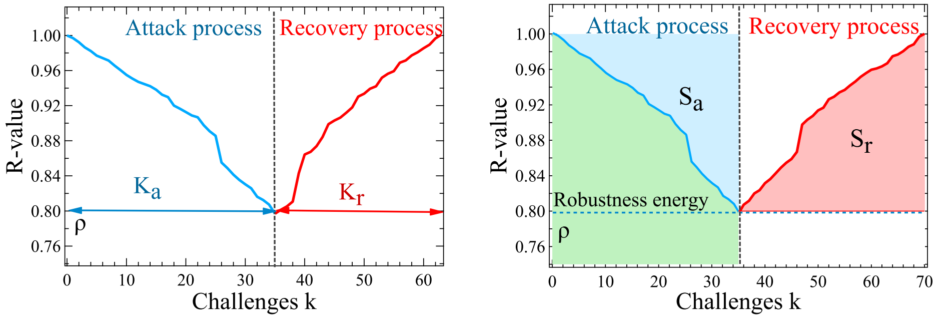
6.2.1. R-VALUE AND CHALLENGE

We inherit the framework and some definitions proposed for network robustness [17, 131] and extend the methodology for the network recoverability. A given network determined by a service and an underlying topology is translated into a mathematical object, defined as the R-value, on which computations can be performed [17]. The R-value takes the service into account and is normalized to the interval $[0,1]$. Thus, $R = 1$ reflects complete functionality in an unattacked network, and $R = 0$ corresponds to absence of performance in a completely destroyed network.

A challenge is an event that changes the network and thus changes the R-value. We assume that a sequence of elementary changes do not coincide in time. Here, we confine an elementary challenge to a link removal in an attack process or a link addition in a recovery process. Since every perturbation has an associated R-value, any realization consisting of a number M of elementary challenges can be described by a sequence of R-values denoted $\{R[k]\}_{1\leq k\leq M}$, where k is the sequence number of challenges.

6.2.2. SCENARIO A: RECOVERY OF ANY ALTERNATIVE LINK

We define R_G as the robustness metric of the network $G(N,L)$ with N nodes and L links. Attacks on a network only consist of link removals in the network by a determined strategy, which usually degrades the robustness of the network. We remove links, one by one, until the R-value R_{G_a} first reaches or drops below a constant ρ, where $\rho \in [0,1]$ is a prescribed R-threshold for the robustness metric that can be tolerated [17]. The above



(a) Scenario A

(b) Scenario B

Figure 6.1: Illustration of the attack process and the recovery process in an Erdős-Rényi (ER) random graph $G_{0.1}(100)$ with link density $p = 0.1$ and network size $N = 100$ in one realization. The R-threshold is $\rho = 0.8$.

process is called the attack or failure process. The number of removed links in the attack process, i.e., *attack challenges*, is denoted by K_a.

Then we launch the recovery process from the remaining network $G_a(N, L - K_a)$. Scenario A assumes that the recovery links can be established between any two nodes in the complement of the graph after attacks. The process of one realization is illustrated in Figure 6.1a. Specifically, we recover the network by adding links, one by one, to the remaining network G_a by a recovery strategy until the robustness metric R-value first reaches or excesses $R_{G_r} = 1$. The network after the recovery process is denoted by $G_r(N, L - K_a + K_r)$, where K_r is the number of *recovery challenges* (adding K_r links). We define the *Link Ratio* denoted η_L as the ratio of the number of attack challenges and the recovery challenges, i.e.,

$$\eta_L(G,\rho) = \frac{K_a}{K_r} \tag{6.1}$$

which indicates the efficiency of the recovery process in one realization. A Link Ratio $\eta_L(G,\rho) > 1$ implies that the network can be recovered by less challenges than the number K_a of attack challenges. Otherwise, the network is more difficult to recover than to destroy.

Scenario A can characterize the recovery process in a connection oriented network with logical connections [132], e.g., a virtual circuit for transporting data or a wireless backhaul network, where the links in a logical network represent the duplex channel between two devices. After such networks are attacked by denial-of-service attacks (DoS) or signal blocking, one should establish several connections or reconfigure several new channels to maintain the network's overall performance. In this case, the overhead cost of the recovery measures mainly depends on the total number of dispatched connections, which corresponds to the number K_r of recovery challenges in Scenario A.

6.2.3. SCENARIO B: RECOVERY OF ATTACKED LINKS

The attack process in Scenario B is the same as in Scenario A. In the recovery process in Scenario B, we add all the links which are removed in the attack process, one by one, until the underlying topology returns to the original. Scenario B describes recovery processes in the physical communication network, e.g., optical backbone networks and power grids supplying to communication networks. In these networks, the recovery measure for each connection, e.g., repairing fiber optic cables, usually requires a relatively long period. During the recovery process, the network still provides services with a degraded performance. Thus, the network recoverability is related to the network performance (or the robustness metric) throughout the recovery process.

One realization of the attack and recovery process is illustrated in Figure 6.1b. In Scenario B, the number of attack challenges and the number of recovery challenges are the same, i.e., $K_a = K_r$, and hence, $\eta_L = 1$ in (6.1). Therefore, we propose another recoverability indicator in Scenario B. The *robustness energy* $S(G,\rho)$ of a network G is the sum of the R-value in the attack process $S(G,\rho) = \sum_{k=0}^{K_a} R[k]$, which expresses the robustness performance of network under successive attacks [131]. Thus, the energy of attack challenges is computed by $S_a(G,\rho) = \sum_{k=0}^{K_a}(1 - R[k]))$, which indicates the cumulative degradation of network performance due to the attacks. In the recovery process, the energy of recovery challenges $S_r(G,\rho) = \sum_{k=0}^{K_a}(R[k] - \rho)$ represents the benefit of the network performance by the recovery measures. The *Energy Ratio* denoted η_E in Scenario B is defined as the ratio between the energy of recovery challenges S_r and the energy of attack challenges S_a in each realization for a determined R-threshold ρ, which follows

$$\eta_E(G,\rho) = \frac{S_r}{S_a} \tag{6.2}$$

An Energy Ratio $\eta_E(G,\rho) > 1$ implies the benefit of recovery measures can compensate the loss of network performance by the attacks, which indicates a high network recovery capability. Conversely, an Energy Ratio $\eta_E(G,\rho) < 1$ implies a low recoverability.

6.2.4. COMPARISON VIA ENVELOPES AND THE RECOVERABILITY INDICATORS

Any realization of attack and recovery processes can be expressed as a sequence of R-values denoted $\{R[k]\}$. To investigate the recovery behavior, we need to know how many challenges k are needed to make R-value decrease to a predefined threshold ρ and increase to its original value, which confines us to investigate the number of challenges K as a function of a specific R-value r, i.e., $\{K[r]\}$. Thus, each value in $\{K[r]\}$ is the number of challenges that is needed to change R-value to a specific R-value r for each realization. The *envelope* is constructed using all sequences $\{K[r]\}$ for $r \in \{r_1, r_2, \ldots, r_H\}$, where $r_j = \rho + \frac{j(1-\rho)}{H}$ is a sampled value and $H = 1000$ is the total sample number. The boundaries of the envelope are given by the extreme number of challenges K

$$K_{\min}[r] \in \{\min(K[r_1]), \min(K[r_2]) \ldots, \min(K[r_H])\} \tag{6.3}$$

$$K_{\max}[r] \in \{\max(K[r_1]), \max(K[r_2]) \ldots, \max(K[r_H])\} \tag{6.4}$$

which gives the best- and worst-case of robustness metrics for a network after a given number of recovery challenges. The expected number of challenges K leading the topological approach r_j is

$$K_{avg}[r] \in \{E(K[r_1]), E(K[r_2]) \ldots, E(K[r_H])\} \tag{6.5}$$

Since $K[r]$ defines a probability density function (PDF), we are interested in the percentiles of $K[r]$

$$K_{m\%}[r] \in \{K_{m\%}[r_1], K_{m\%}[r_2] \ldots, K_{m\%}[r_H]\} \tag{6.6}$$

where $K_{m\%}[r]$ are the points at which the cumulative distribution of $K[r]$ crosses $\frac{m}{100}$, namely $K_{m\%}[r] = t \Leftrightarrow \Pr[K[r] \leq t] = \frac{m}{100}$.

We apply the envelope to present the behavior of the attack and recovery processes on a network [17, 131]. The envelope profiles a rough PDF of the random variables of the number of challenges K, which is the probability of a random variable to fall within a particular region. The area of the envelope can be regarded as the variation of the robustness impact of a certain series of challenges, which quantifies the uncertainty or the amount of risk due to perturbations.

We propose two recoverability indicators, the Link Ratio $\eta_L(G,\rho)$ and the Energy Ratio $\eta_E(G,\rho)$, for different scenarios, respectively. Since an attack process and a recovery process could be random under the random strategy, the recoverability indicators are random variables. We can compare the recoverability of different networks by the average recoverability indicators for simplicity. For example, the average Link Ratio $E[\eta_L(G_1,\rho)] > E[\eta_L(G_2,\rho)]$ for two different networks G_1 and G_2 implies that the network G_1 usually has a better recoverability than G_2 in Scenario A for the robustness threshold ρ.

Besides the average recoverability indicators, we are also concerned about the variance of the recoverability indicators $Var[\eta(G,\rho)]$, which indicates how likely the recoverability is to shift upon the random strategy. A smaller variance of the recoverability indicators $Var[\eta(G,\rho)]$ implies a narrower uncertainty of the recoverability indicators, thus a better recoverability.

6.3. ROBUSTNESS METRIC AND RECOVERY STRATEGY

In this section, we introduce the main factors of a specific recovery process, which involve robustness metrics, recovery strategies and network topologies.

6.3.1. ROBUSTNESS METRICS

A group of topological metrics are proposed to measure the network robustness [122] and the correlation of some metrics in random graphs and functional brain networks are investigated in [133]. We select 20 real telecommunication networks in the specialized databases [134], and show the correlation between metrics. As metrics, we include the network efficiency E_G, the spectral radius of adjacency matrix λ_1, the algebraic connectivity μ_{N-1}, the diameter φ, the effective graph resistance r_G, the ratio μ_1/μ_{N-1}, the average hopcount $E[H]$ among all node pairs, the clustering coefficient c_G in Figure 6.2. Considering services of communication networks, we select 1) the network

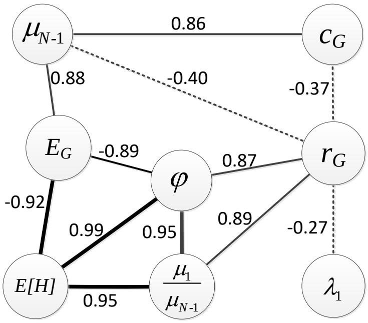


Figure 6.2: Correlation of several topological properties in 20 real telecommunication networks. The Pearson correlation coefficients ρ_{Pearson} are marked if the correlation is strong enough, i.e., $|\rho_{\text{Pearson}}| > 0.8$ (by solid lines) or weak enough, i.e., $|\rho_{\text{Pearson}}| < 0.4$ (by dash lines).

efficiency E_G and 2) the effective graph resistance r_G as the robustness metrics that characterize the performance of end-to-end transmissions. Figure 6.2 shows that these two path-based metrics, the network efficiency E_G and the effective graph resistance r_G, are comparatively lowly correlated (i.e., $\rho_{\text{Pearson}}(E_G, r_G) = -0.63$).

1) **Network efficiency E_G.** We assume that the hopcount $h(i, j)$, i.e., the number of links in the shortest path from node i to j, indicates the overhead of data delivery from end to end. Thus, the reciprocal of the hopcount $1/h(i, j)$ implies the amount of packages for one unit overhead, which can be interpreted as the efficiency of the communication between two nodes. If there is no path from i to j, $h(i, j) = \infty$ and $1/h(i, j) = 0$. For a whole network, the efficiency of a given network can be computed by the mean of the reciprocals of all the hopcount $h(i, j)$ in a network, i.e.,

$$E_G = \frac{\sum_{i \neq j \in G} 1/h(i, j)}{\binom{N}{2}} \tag{6.7}$$

which is defined as network efficiency [130]. Network efficiency quantifies how efficient the exchange of information across the whole network under the shortest-path routing [135]. The network efficiency of a network monotonically decreases with the successive link removals.

2) **Effective graph resistance r_G.** The effective graph resistance [64, 136–138] origins from the field of electric circuit analysis, which is defined as the accumulated effective resistance between all pairs of nodes. The effective graph resistance refers to the average power dissipated in a resistor network with random infected currents, which can indicate the overall diffusivity of information spreading in a communication network. Also, the effective graph resistance r_G determines the onset of congestion in a communication network. Specifically, let δ be the average total input rate of the network. It can be shown [139] that the maximum acceptable value of δ, which ensures that all links are within their transmission capacity, is upper bounded by $\binom{N}{2} r_G^{-1}$.

To generalize the impact of attacks on the robustness metrics, we apply the reciprocal of the effective graph resistance r_G^{-1} as the R-value, which decreases for link removals in an attack process and increases in a recovery process. In this chapter, we also assume that the removed links leading to the network disconnection have the supreme priority and can be restored instantly. Thus, we numerically exclude the realizations of attack processes that disconnect the network and lead to $r_G^{-1} = 0$.

6.3.2. ATTACK AND RECOVERY STRATEGIES

For simplicity and generality, we consider a *random attack* strategy in attack processes. The random attack strategy implies that the attacks or failures occur independently on links randomly and uniformly, which is consistent with the random failure stage in a product life cycle. The R-value $R[k]$ for a determined number of attack challenges k is a random variable. We consider three different strategies for recovery measures, i.e., random recovery, metric-based recovery and greedy recovery:

1) **Random recovery:** The random recovery strategy refers to the strategy that the links are added randomly and uniformly, one by one, in recovery processes, which can describe a self-repairing process after attacks or recovery measures without scheduling.

2) **Metric-based recovery:** The metric-based strategy determines the sequence of adding links by the topological metrics of links. The performance of a network is usually restricted by its structural “bottleneck”, i.e., the effective graph resistance is related to the algebraic connectivity and the minimum degree [133]. A good recovery strategy tends to remedy such bottleneck. Thus, we consider two metrics of links between node i and j: the minimum product of degree $d_i d_j$, and the minimum product of eigenvector centrality $c_i c_j$. For each challenge in a recovery process, we select and restore the link l_{ij}^* with the minimum $d_i d_j$ or $c_i c_j$.

3) **Greedy recovery:** The greedy recovery strategy involves adding the link l^* that makes R-value increase most in each challenges, i.e.,

$$l^* = \arg\max_{l \in G^c} R(G+l) - R(G) \tag{6.8}$$

where G^c is the complement of the current network G. The greedy strategy is a practical and intuitive recovery strategy, where the current optimal link for improving the performance of the network has the priority to be recovered during a recovery process.

6.3.3. TELECOMMUNICATION NETWORKS

We select 10 real communication networks for case study. The topological properties of the 10 real telecommunication networks are described in Table 6.1. These 10 telecommunication networks include logical networks (representing the IP layer) and backbone transport networks (connected with optical fiber). This set of networks was selected in specialized databases [134], covering the telecommunication systems located in different areas of the world, i.e., DFN (German backbone X-WiN network), Cernet (China education and research network), Bt_US (Internet provided by BT in the US), GtsCe (GTS network in central Europe), Cogentco (IP backbone network provided by Cogentco), TataNld (Tata national long distance network), ATT_US (IP MPLS backbone network provided by AT&T), Coronet (IP backbone network provided by Cogent), GEANT (IP backbone network provided by GEANT), Renater (Internet provided by Renater).

Networks	N	L	$E[D]$	λ_1	μ_{N-1}	φ	ρ_D	E_G
DFN	58	87	3.00	5.43	0.25	6	-0.11	0.36
Cernet	41	58	2.83	4.78	0.22	5	-0.35	0.40
Bt_US	36	76	4.22	5.85	0.41	6	0.03	0.44
GtsCe	149	193	2.59	3.81	0.01	21	-0.09	0.16
Cogentco	197	243	2.47	3.79	0.01	28	0.03	0.14
TataNld	145	186	2.57	3.27	0.01	28	-0.21	0.15
ATT_US	25	56	4.48	5.76	0.65	5	-0.02	0.51
Coronet	100	136	2.72	3.30	0.05	15	0.04	0.20
GEANT	40	61	3.05	4.32	0.14	8	-0.20	0.36
Renater	43	56	2.60	3.88	0.10	9	-0.15	0.33

Table 6.1: Topological properties, explained in Section 6.3.1, of the 10 real telecommunication networks. The degree assortativity is denoted by ρ_D.

6.4. RESULTS AND DISCUSSION

In this section, detailed results and analysis on the real-world network via the proposed approach for assessing network recoverability are presented. For some evaluation items, we only present results for a specific network, i.e., DFN. We set the R-threshold as $\rho = 0.8$ in the following simulations. The approach translates easily to other networks or other robustness metrics.

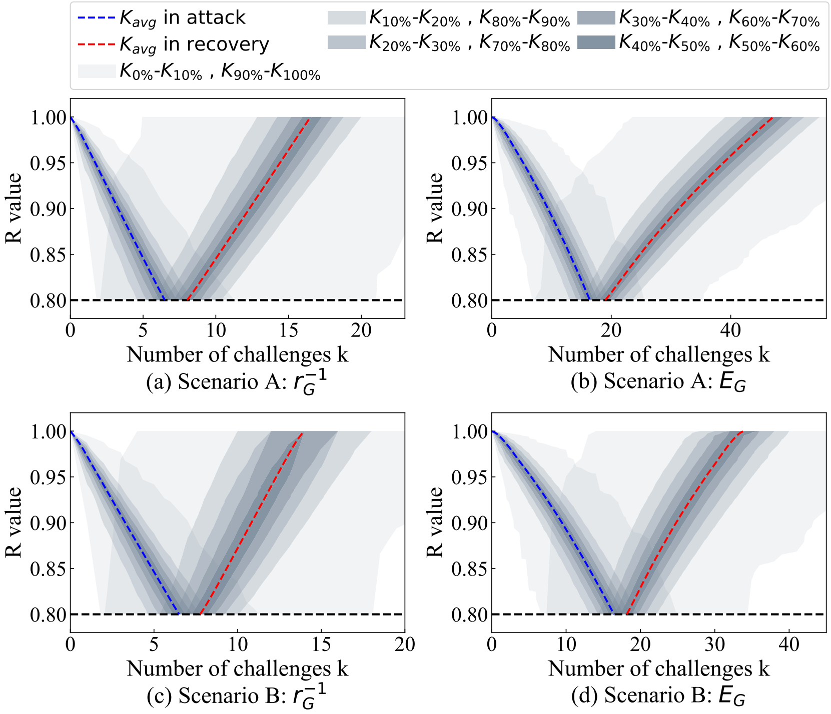
6.4.1. ENVELOPE EXAMPLES AND COMPARISON

Each realization of processes consists of an attack process and a following recovery process. Figure 6.3 exemplifies the envelopes of the challenges in DFN network for two scenarios and two robustness metrics, r_G^{-1} and E_G, respectively, under the random recovery strategy. The envelopes for the attack processes are similar in different scenarios while Scenario A usually needs more challenges to recover the robustness metrics than Scenario B, if the random recovery strategy is employed. The total number of challenges $K_a + K_r$ could cover a wide range of values since the number of challenges $K_a + K_r$ is influenced by two random processes (i.e., attack and recovery).

Figure 6.3a and Figure 6.3c also illustrates that the function R-value of the average number of challenges $R[K_{avg}]$ for the robustness metric r_G^{-1} is almost linear, in both the attack process and the recovery process. For the robustness metric E_G, the function $R[K_{avg}]$ is slightly concave, illustrated in Figure 6.3b and Figure 6.3d. We will show that the concavity of the function $R[K_{avg}]$ could help to explain the behavior of the recoverability indicators.

6.4.2. COMPARISON OF RECOVERY STRATEGIES

The envelope computation can be applied to compare the performance of different recovery strategies for a specific realization of attacks. Figure 6.4 shows different recovery strategies (e.g., random, minimum $d_i d_j$, minimum $c_i c_j$, greedy) for one realization of attack processes under random attack strategy in DFN network. The envelope of recovery processes by random recovery for the network efficiency E_G covers a larger surface than that of the inverse of the effective graph resistance r_G^{-1}. This implies that

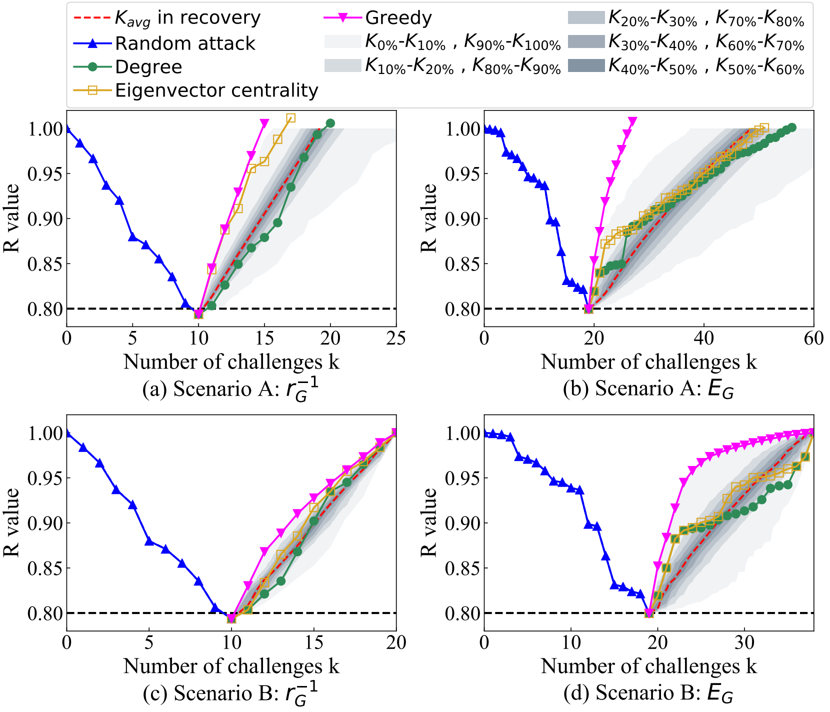


(a) Scenario A: r_G^{-1}
(b) Scenario A: E_G
(c) Scenario B: r_G^{-1}
(d) Scenario B: E_G

Figure 6.3: Envelopes of the challenges for two scenarios and two robustness metrics (i.e., the inverse of the effective graph resistance r_G^{-1} and the network efficiency E_G) in DFN network, by random recovery strategy. Each envelope is based on 10^4 realizations.

the network efficiency E_G in different realizations could deviate more from one another under the random recovery, and the performance of random recovery is more difficult to be guaranteed. The average challenge sequence $\{K_{avg}\}$ under the random recovery can be a standard to evaluate the performance of other recovery strategies.

Figure 6.4 shows that the performance of metric-based strategies, e.g., minimum degree product and minimum eigenvector centrality product, is not guaranteed. Especially for the network efficiency E_G, the metric-based strategies outperform the average of random strategy in the initial stage of recovery processes but degrade for more recovery challenges. Meanwhile, we notice that the greedy recovery usually upper bounds the random recovery envelopes. The R-value as a function of the number of challenges k under the greedy strategy is concave in the recovery process, which demonstrates the diminishing returns property of the recovery measures. Since the optimal recovery strategy is usually an NP-hard problem, we suspect that the greedy recovery can be a practical near-optimal recovery strategy for both robustness metrics, i.e., r_G^{-1} and E_G.



(a) Scenario A: r_G^{-1} (b) Scenario A: E_G

(c) Scenario B: r_G^{-1} (d) Scenario B: E_G

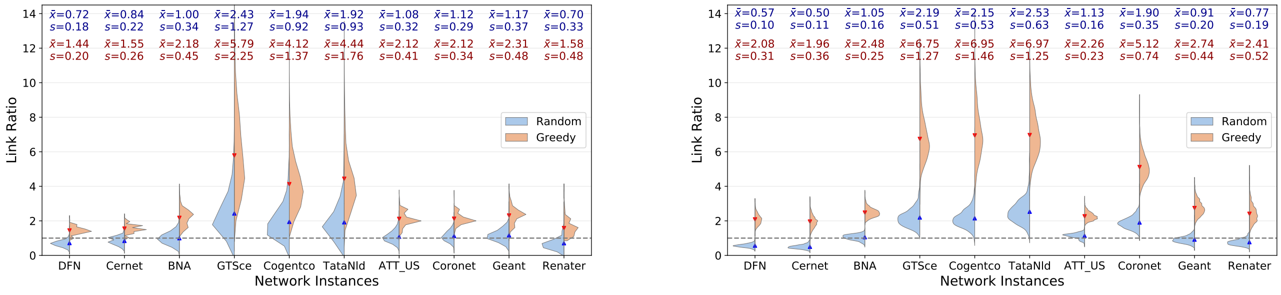
Figure 6.4: Comparisons of different recovery strategies for one realization of attacks in DFN network. Two scenarios and two robustness metrics (i.e., the inverse of the effective graph resistance r_G^{-1} and the network efficiency E_G) are applied. Each envelope is based on 10^4 realizations.

6.4.3. OVERVIEW OF THE LINK RATIO AND THE ENERGY RATIO

We employ the proposed approach and the recoverability indicators η (including the Link Ratio η_L and the Energy Ratio η_E) to evaluate the 10 real telecommunication networks. Figure 6.5 shows the recoverability indicators under two different scenarios, two robustness metrics and two recovery strategies for 10 networks by violin plots. Violin plots are similar to box plots, except that they show the probability density of the ratios η at different values, which presents more insights about the ratios η under random circumstances. Moreover, violin plots can be applied to compare the performance of any two different strategies, e.g., the random and the greedy.

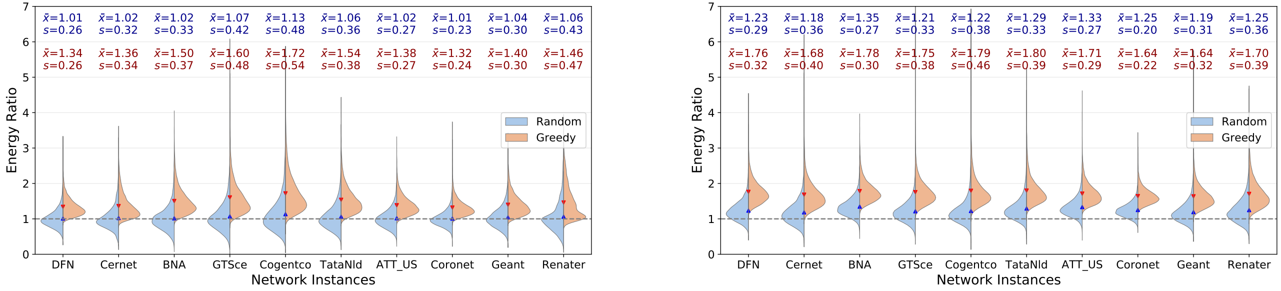
Figure 6.5 shows that almost all histograms of the ratio η, regardless of the scenarios, the strategies and the metrics, exhibit heavy-tailed distributions, while the greedy strategy presents a longer tail. Also, the ratio η has a wider range of values under the greedy strategy, which implies the greedy strategy has a higher probability to lead to a large ratio η, as well as a better recovery performance.

For both robustness metrics in Scenario A, DFN, Cernet and Renater have an average Link Ratio $E[\eta_L] < 1$ for the random strategy, which implies a relatively low recovery capability. By contrast, GTSce, Cogentco and TataNld have a large average Link Ratio $E[\eta_L] > 1$, which outperform other networks much under both the random strategy and



(a) Scenario A: r_G^{-1}

(b) Scenario A: E_G



(c) Scenario B: r_G^{-1}

(d) Scenario B: E_G

Figure 6.5: Violin plots of the Link Ratio η_L in Scenario A and the Energy Ratio η_E in Scenario B. The average ratios $\tilde{x} = E[\eta]$ and the standard deviations $s = \sqrt{Var[\eta]}$ are presented on the top of each subplot. The blue surface and blue notes represent the random recovery strategy, and the red surface and blue notes represent the greedy recovery strategy. The average ratios are marked as triangle markers. Each histogram of η is based on 10^4 realizations.

the greedy strategy. The network with a larger average Link Ratio usually has a larger diameter φ, then the new established links in Scenario A could shorten the diameter φ and increase the topological approach (r_G^{-1} or E_G) more.

The Energy Ratio η_E presents different behaviors of the Link Ratio η_L compared with Scenario A. The average Energy Ratios $E[\eta_E]$ for the robustness metric r_G^{-1} approximate 1 under the random strategy, which can be explained by the fact that the function $R[K_{avg}]$ is almost linear (illustrated in Section 6.4.1), and thus the energy $S_a \approx S_r$. Since the function $R[K_{avg}]$ is concave for the robustness metric E_G^{-1} and thus the energy $S_a < S_r$, the average Energy Ratios $E[\eta_E]$ for different networks are slightly larger than 1. The average Energy Ratio $E[\eta_E]$ in Scenario B under the greedy strategy is usually located in the tail of the distribution of the Link Ratio η_L under the random strategy, which demonstrates that the greedy strategy can update the recoverability of networks much.

6.4.4. RELATION BETWEEN SCENARIO A AND SCENARIO B

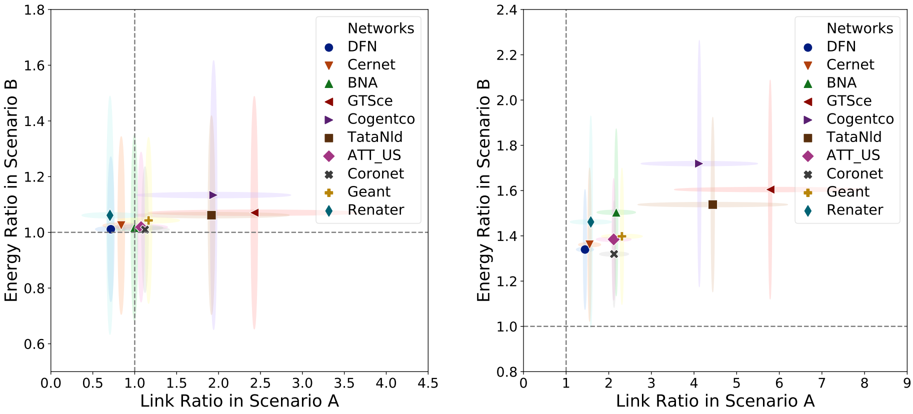
To compare the recoverability among different networks, we employ the SA-SB plots to show the relation of the Link Ratio in Scenario A and the Energy Ratio in Scenario B under a determined recovery strategy. SA-SB plots are divided as 4 quadrants by the reference lines of the Link Ratio $\eta_L = 1$ and the Energy Ratio $\eta_E = 1$ in order to easily assess the recoverability by the location of the average ratios $E[\eta_L]$ and $E[\eta_E]$. Figure 6.6 shows the average ratios $E[\eta]$ and the standard deviations $\sqrt{Var[\eta]}$ for the real-world networks in SA-SB plots.

Figure 6.6 shows that the recoverability under two different scenarios has a weak correlation, e.g., a Link Ratio η_L in Scenario A does not lead to a higher η_L in Scenario B. We can also observe that all the average Energy Ratios $E[\eta_E]$ are located in the first and the second quadrant, which demonstrates a good recoverability of tested networks in Scenario B. However, the average Link Ratios $E[\eta_L]$ in the second quadrant suggest the topological improvement for these networks in Scenario A.

For a determined robustness metric, both the average Link Ratio $E[\eta_L]$ and the Energy Ratio $E[\eta_E]$ can be increased by applying the greedy strategy, but the performance can be different. For example, the average Link Ratio $E[\eta_L]$ of Cogentco is larger than that of TataNld under the random strategy but smaller than that of TataNld under the greedy strategy, which implies that the performance of a recovery strategy strongly depends on the network topology.

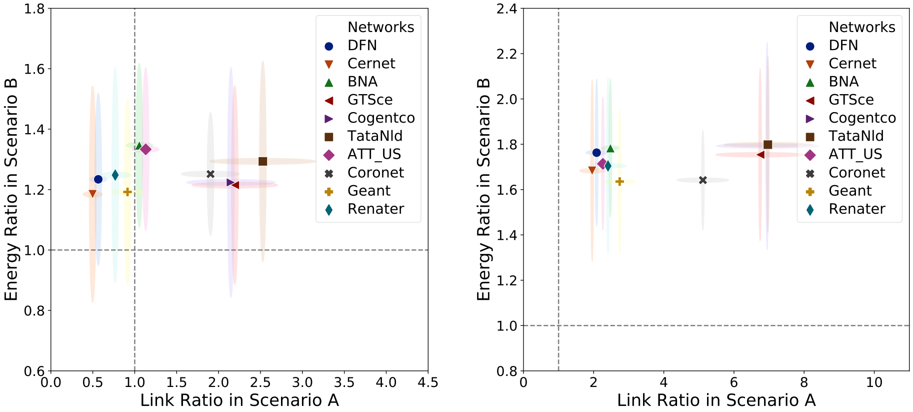
6.4.5. IMPACT OF R-THRESHOLD

Figure 6.7 exemplifies that a larger R-threshold ρ slightly decreases the average Energy Ratio $E[\eta_E]$ in Scenario B, while the average Link Ratio $E[\eta_L]$ in Scenario A decreases more with a larger R-threshold ρ (i.e., a lower damage level). This result implies that the marginal cost (i.e., the number of recovery challenges to increase a same fraction of R-value) decreases with a higher damage level, which is in line with the concavity of the function $R[K_{avg}]$ under the random strategy (illustrated in Figure 6.3) and the function $R[k]$ under the greedy strategy (illustrated in Figure 6.4). In these cases, one challenge of recovery can increase R-value more for a smaller R-value.



(a) Random recovery: r_G^{-1}

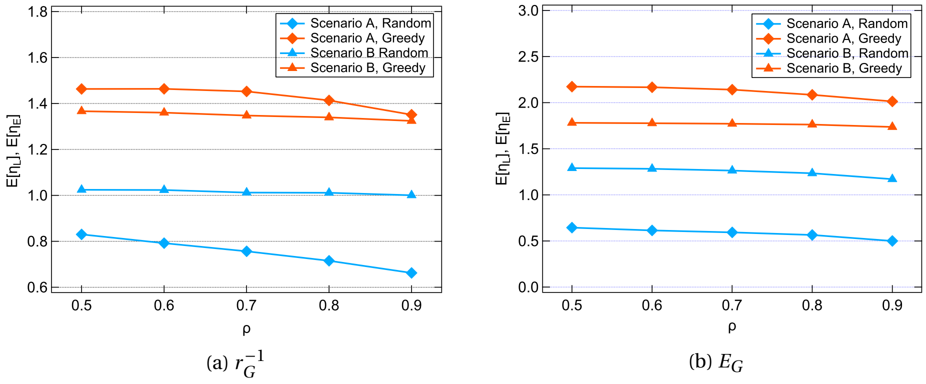
(b) Greedy recovery: r_G^{-1}



(c) Random recovery: E_G

(d) Greedy recovery: E_G

Figure 6.6: SA-SB plots of the Link Ratio η_L and the Energy Ratio η_E for two robustness metrics (i.e., the inverse of the effective graph resistance r_G^{-1} and the network efficiency E_G). The dark markers represent the average ratios $E[\eta]$, and the cross indicates the value range $\left[E[\eta]-\sqrt{Var[\eta]}, E[\eta]+\sqrt{Var[\eta]}\right]$.



(a) r_G^{-1} (b) E_G

Figure 6.7: Average Link Ratio $E[\eta_L]$ and average Energy Ratio $E[\eta_E]$ as a function of the robustness threshold ρ in DFN network.

6.5. CHAPTER SUMMARY

This chapter proposes a topological approach for evaluating the network recoverability in two scenarios, which extends the application of the framework [17] of network robustness. We assess the recoverability of 10 real communication networks for two different path-based robustness metrics, i.e., the network efficiency and the effective graph resistance. In accordance with the results, the network recoverability presents different behaviors between link-based Scenario A and energy-based Scenario B. All the telecommunication networks have a healthy recovery capability in Scenario B under the random recovery strategy, i.e. the average Energy Ratio $E[\eta_E] > 1$, while three of the networks (DFN, Cernet and Renater) suggest topological improvements for the recoverability in Scenario A, i.e., the average Link Ratio $E[\eta_L] < 1$. The goodness of the recoverability in Scenario B can be explained by the concavity of the R-value as a function of the number of challenges. The network recoverability is also strongly related to the recovery strategy. Comparing the performance of different recovery strategies, the greedy recovery strategy exhibits a good performance for the investigated robustness metrics and thus improves the network recoverability.

7

ROBUSTNESS ASSESSMENT OF MULTIMODAL FREIGHT TRANSPORT NETWORK

Multimodal freight transport allows switching among different modes of transport to utilize transport facilities more efficiently. This chapter proposes an approach on network modeling and robustness assessment for multimodal freight transport networks, where the nodes represent junctions, terminals and crossings, and the links represent pathways. The network model captures the features of interconnection and interdependency. Freight can switch between different modalities at interconnected terminals, while disruption of a single interdependent node (e.g., bridge, tunnel, railway crossing) affects multiple modalities. Considering disruptions of infrastructure elements and capacity degradation of pathways as perturbations, the network robustness is evaluated as the increment of the total travel time caused by these perturbations. We apply our robustness assessment model to the Dutch freight transport, taking into account three modalities: inland waterway, road and railway. The node criticality, defined as the impact of a node removal on the total travel time, resembles a power-law distribution, independent of different traffic assignments. This scale-free property implies a relatively robust state of the network against single random disruptions. Further, we show that the most critical nodes can be roughly identified by their topological properties. Our research can help to schedule the maintenance of the infrastructure by assigning priority to the critical infrastructure.

7.1. INTRODUCTION

THE European hinterland freight transport sector has aimed at a shift towards multimodal transport such as railway, inland waterway and sea transport, in order to alleviate the saturation of road systems [140][141]. Intermodal transport and synchromodal transport are promoted as two promising solutions. Intermodal freight transport allows moving goods by using various modalities consecutively [142]. Synchromodal transport aims at real-time and flexible switching among different modes of transport according to the latest logistic information, so as to utilize all transport facilities more efficiently [143][144].

Transport networks are prone to suffer from various perturbations, for example infrastructure failures and temporary closures due to construction work. More severe perturbations, such as strikes or extreme weather (e.g. droughts, heavy snow, strong winds), could lead to the partial unavailability of transport networks. Network robustness is interpreted as a measure of the network's response to perturbations or challenges imposed on the network [145] [146], which considers both the network topology and the service for which the network is designed [17]. The robustness of transport networks has been studied extensively in recent years, mainly based on two variants of perturbations: connectivity related and capacity related. Connectivity related perturbations [147–149] regard the failure of infrastructure elements as removed nodes/links from a network. Capacity-related perturbations [150–153] consider failures of infrastructure elements as the capacity reduction of parts of the network, as opposed to a complete removal of parts of the network. Despite that the above assessment approaches have been applied in different modes of transport, the framework of robustness assessment for multimodal networks has seldom been studied [154]. The issues including (i) how to abstract particular infrastructure (e.g. terminals and crossings) of multimodal transport? (ii) how to assess the network robustness under structural disruptions? (iii) how do different traffic assignments impact the robustness performance? are still open. This study aims at addressing the above questions.

Introducing concepts of *Network Science* to transportation research allows us to propose a framework of network modeling, robustness assessment and critical structure identification for multimodal freight transport. An interdependent network in Network Science is a multi-layer network consisting of different types of networks that depend upon each other for their functioning [155]. The interdependency in networks has been applied to measure the robustness of communication networks, that control and are supported by power grids [156], notably by investigating the impact of cascading failure [157][158]. The representation of interdependent networks is an excellent proxy for the structure of multimodal transportation systems. Taking into account several modalities (e.g. inland waterway, road and railway), the transport infrastructure is modeled as a multilayered network, where the nodes represent junctions, terminals and crossings, and the links represent pathways. This network features two properties: interconnection [159] and interdependency [157]. Specifically, transloading terminals are facilities where freight can be transferred from one mode of transport to another and are modeled as interconnections. The crossings, whose functioning influences multiple modes of transport (e.g. bridges), are interdependent nodes. Thus, the disruption of a crossing implies a simultaneous unavailability of related pathways in multiple layers of the network. Consequently, this macroscopic network model abstracts the intricate connectivity of multimodal transportation networks, and also characterizes various types of network perturbations.

Based on the network model, we assess the robustness of multimodal transportation networks. We regard the total travel time of transporting all the freights as a performance indicator, which usually increases due to disruption of any infrastructure element [160]. Our framework assumes that the increment of total travel time due to a node removal reflects the criticality of this node. Then, the robustness of a network can be measured by considering the time increment arising from every node removal in a statistical way. Further, we explore the correlation between the time increment due to a node removal and the topological properties of this node, which helps to identify the critical nodes faster.

We assess our approach by an extensive case study on the freight transport network in the Netherlands. The case study is not limited to the traffic assignment of all-or-nothing (AoN) [161], but also more practical traffic models, including modal split (MS) [162], user equilibrium (UE) [163] and system optimum (SO) [163]. We investigate the robustness performance and topological properties of the critical nodes under different traffic assignments. The assessment under single element disruptions identifies the critical nodes, whose disruption leads to a relatively high increment of the travel time. The critical nodes need to be given a higher priority for repairs and maintenance by the responsible organization. The robustness assessment under the capacity degradation of pathways can help to evaluate the impact of a large-scale disaster and work out contingency plans. A general recovery framework for any type of network is presented in [117], which allows to assess the performance of recovery measures in transport networks.

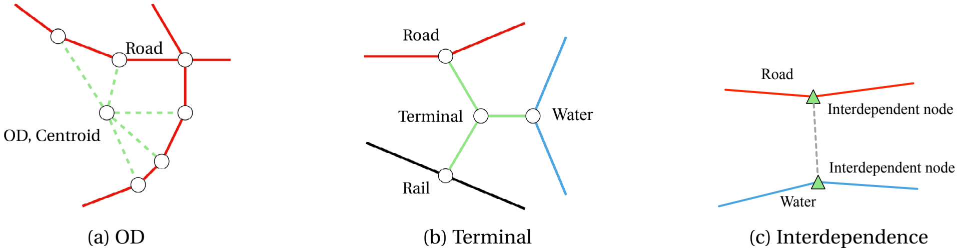
The main contribution of this chapter can be summarized as:

1. We introduce the concepts of interconnection and interdependency into modeling multimodal networks, which fill the gap for evaluating the impact due to disruptions of the terminals and the crossings.
2. The framework of transport network assessment comprehensively considers both disruptions/degradation of infrastructure and different traffic behavior.
3. The assessment framework develops the method of roughly identifying critical nodes by nodal topological properties.

This chapter is organized as follows. We introduce the method for modeling a multimodal transport network in Section 7.2. Section 7.3 proposes a framework of robustness assessment and defines a robustness indicator. We apply the assessment method to the Dutch freight transport network in Section 7.4. Section 7.5 summarizes our findings and concludes this chapter.

7.2. THE SYNCHROMODAL NETWORK MODEL

In this section, we briefly introduce the network model for multimodal transport, which features multiple layers, interconnection and interdependency.



(a) OD (b) Terminal (c) Interdependence

Figure 7.1: Illustration of network modeling. (a) Connections between OD centroid and main roads. Solid lines represent main roads, and green lines represent small roads. (b) Connections between terminal and main roads of modalities. Green lines represent small roads. (c) Relation between two modalities with an interdependent node.

7.2.1. MULTI-LAYER NETWORK

The underlying topology of the multimodal freight transportation can be represented by an undirected network $G(\mathcal{N},\mathcal{L})$ with the set $\mathcal{N}$ of N nodes and the set $\mathcal{L}$ of L links. The nodes in the network represent transloading terminals, crossings and junctions, which are connected by the links as different types of pathways. In our network model (as illustrated in Figure 7.1), we consider three modalities: road (Road), railway (Rail) and inland waterway (Water), which are the most common modalities in many European countries. The underlying topology of each modality is represented by a subgraph $G_m(\mathcal{N}_m,\mathcal{L}_m)$ for modality $m \in \{Road, Rail, Water\}$. The terminals providing the interconnection between different modalities are represented by the subgraph G_T, and the origins and destinations (OD) locating the sources and targets of cargo demands are defined in the subgraph G_{OD}. The whole transport network is a supergraph which consists of multiple undirected subgraphs. Thus, the resulting supergraph combining all infrastructure elements is defined as $G(\mathcal{N},\mathcal{L})$ with node and link set,

$$\mathcal{N} = \mathcal{N}_{Road} \cup \mathcal{N}_{Rail} \cup \mathcal{N}_{Water} \cup \mathcal{N}_T \cup \mathcal{N}_{OD} \tag{7.1}$$

$$\mathcal{L} = \mathcal{L}_{Road} \cup \mathcal{L}_{Rail} \cup \mathcal{L}_{Water} \cup \mathcal{L}_T \cup \mathcal{L}_{OD} \tag{7.2}$$

7.2.2. ORIGINS AND DESTINATIONS (OD CENTROID)

Centroids of regions are used [154] [164] to model the origins and destinations in each region, where the centroid of a polygonal area is located at the center of mass of that polygon. We split the country into several regions, where the centroid of each region represents the origin and destination of the freight transport demand in that region. These centroids are represented by the nodes in the subnetwork G_{OD}. We assume that the centroid is connected to all access points in the road subgraph of this region by small roads. The small roads are represented by the OD links in the subnetwork G_{OD}. Figure 7.1a illustrates the network model for centroids.

The amount of freight (in tons) that is transported between the origins and destinations is defined in the demand matrix D, where the element D_{ij} defines the average amount of cargo transported from region i to region j. Freight transported between the origin and the destination in the same region is not considered, and thus the diagonal elements of the demand matrix D are all equal to 0.

7.2.3. INTERCONNECTION

Intermodal transport allows switching among different modes of transport at transloading terminals. At a transloading terminal two or more modalities are interconnected. Figure 7.1b illustrates the network model for these terminals. Each transloading terminal is represented by a node in the subgraph G_T. The terminal node is connected with links to nodes of the appropriate modes of transport. This modeling method captures the feature of disruptions of terminals, i.e. the freight cannot switch to a different mode of transport if the terminal node is removed.

7.2.4. INTERDEPENDENCY

The three modalities (waterway, road and rail) cross each other regularly (see Figure 7.3 below). At each crossing, a civil engineering structure is needed (e.g. bridge, tunnel, railway crossing) to efficiently use both modalities. Thus, the disruption of a single civil engineering structure can affect multiple modalities, which causes the interdependency between the modalities. For example, the disruption of a bridge can affect both the road and the waterway simultaneously. Unlike for interconnection, transloading of freight is not possible at these crossings. Figure 7.1c illustrates the network model of an interdependent node. We represent each crossing structure as two nodes in the subgraphs of the two modalities. The relation of interdependency between these two nodes implies a simultaneous removal due to a disruption, i.e. if either node is removed, the other node will be removed as well.

7.3. NETWORK ROBUSTNESS ASSESSMENT

Our robustness assessment will be described in detail in this section based upon Figure 7.2. The proposed framework encompasses two aspects: the infrastructure designed by network operators and traffic behavior determined by drivers, which corresponds to the fact that the network robustness is related to its underlying topology and services [17]. For the transport network, we can compute the predefined performance indicator of the fully-functional network based on specific OD-demands and a given traffic assignment. Both the degradation of the functionality of pathways and the disruptions of transport elements could change the network performance indicator (i.e. the total travel time). For the same OD demands and the same traffic assignment, we recompute the performance indicator in the perturbed network. We then assess the robustness of the whole network based on the changes of the performance indicator due to the perturbations. The framework provides the method for exploring the relation between the impact of the disrupted elements on the performance and several topological properties of these elements.

7.3.1. LINK ATTRIBUTES

The performance of the transport system is related to the route condition of each modality, which translates to the link attributes in the network model. Each pathway

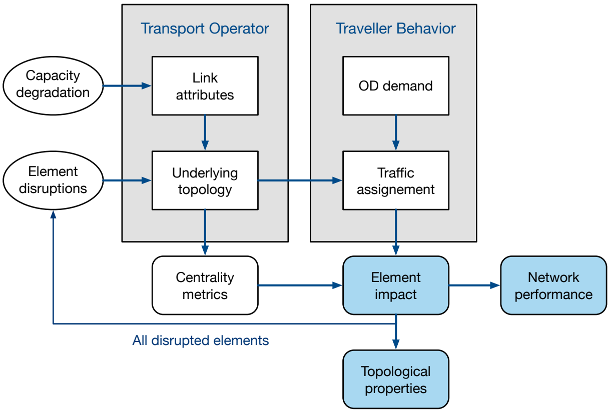

Transport Operator
Traveller Behavior
Capacity degradation
Link attributes
OD demand
Element disruptions
Underlying topology
Traffic assignement
Centrality metrics
Element impact
Network performance
All disrupted elements
Topological properties

Figure 7.2: Framework of transport network assessment.

segment represented by a link ℓ in the network has two attributes: the free-flow average speed v_ℓ (in kilometers/hour) and the capacity c_ℓ (in tons/hour). The average speed v_ℓ (in kilometers/hour) determines the travel time t_ℓ (in hours) on link ℓ, i.e. $t_\ell = \frac{d_\ell}{v_\ell}$, where d_ℓ is the length of link ℓ (in kilometers). Considering the travel time as the cost of freight transportation, the travel time t_ℓ is regarded as the link weight in the network for computing the shortest path from the origin to the destination.

The average speed v_ℓ is a constant for the traffic assignments without capacity constraint, i.e. the link travel time t_ℓ is also constant for any amount of flow x_ℓ (in tons/hour) assigned on the link ℓ. In contrast, the link capacity constraints are included in the travel time functions considering traffic congestion. A popular form of these functions that reflects the travel time $t_\ell(x_\ell)$ for each vehicle as a function of the flow x_ℓ (in tons/hour) on link ℓ, proposed by the Bureau of Public Roads (BPR), is given by [161]:

$$t_\ell(x_\ell) = t_{0\ell}\left(1 + \alpha\left(\frac{x_\ell}{c_\ell}\right)^\gamma\right) \tag{7.3}$$

where c_ℓ is the capacity of link ℓ (in tons/hour), and $t_{0\ell} = \frac{d_\ell}{v_\ell}$ is the free flow travel time, α and γ are the the shape coefficients, for which the value of $\alpha = 0.15$ and $\gamma = 4$ are generally applied [161]. The travel time t_ℓ without considering congestion is consistent with an infinite link capacity $c_\ell \to \infty$.

7.3.2. NETWORK PERTURBATION

Network perturbations refer [117][131] to two scenarios: the disruptions of transport elements and the degradation of capacity, which translate to the changes in the underlying topology and the link attributes, respectively. For the scenario of node disruptions, the unavailable nodes are regarded to be removed from the original underlying topology. In this chapter, we mainly consider disruptions of individual nodes to investigate the impact of each node on the performance of the network. This analysis may help to identify the critical infrastructure elements and schedule improvement measures. The scenario of node disruptions also includes the case that multiple nodes can be removed simultaneously, which describes simultaneous accidents or cascading failures [157]. We define a *random failure* as the failure scenario where a given fraction of multiple elements are removed from the network uniformly at random [117].

The scenario of capacity degradation refers to the effect of capacity reduction of a single modality on the performance of the whole network. This scenario aims to describe the cases of large-scale natural and man-made disasters, e,g. driver strikes and low water levels in rivers due to droughts.

7.3.3. PERFORMANCE INDICATOR

We denote by z_ℓ the total amount of freight (in tons) attempting to use link ℓ as the flow rate x_ℓ (in tons/hour) within h hours, i.e., $z_\ell = x_\ell h$. Invoking that the delay time for each vehicle traveling on link ℓ is t_ℓ specified by (7.3), we apply the total travel time, also called the total delay time, C_G (in tons·hours) of transporting all the freights among all links to measure the performance of network G, which is defined as

$$C_G = \sum_{\ell \in \mathcal{L}} z_\ell t_\ell (x_\ell) \tag{7.4}$$

The network performance indicator C_G usually increases to $C_G + \Delta C_G$ due to node disruptions and degradation of link attributes. The robustness can be measured by the normalized increment of the total travel time due to the perturbation on the network, which defines the robustness indicator η_G as

$$\eta_G = \frac{C_{G'} - C_G}{C_G} = \frac{\Delta C_G}{C_G} \tag{7.5}$$

where G' is the network after perturbations.

For measuring the impact of a single node disruption on the performance of the network, we define the *node criticality* as the normalized increment of the total time caused by the node removal. The node criticality of node i is the same as the robustness indicator η_G under the single node disruption, which is

$$\omega_i = \frac{C_{G \backslash \{i\}} - C_G}{C_G} \tag{7.6}$$

where $G \backslash \{i\}$ is the graph in which node i is removed as well as all its incident links. The network robustness under isolated single disruptions can be measured by the average node criticality $E[\omega] = \frac{1}{N}\sum_{i=1}^{N} \omega_i$ among all nodes. In general, a smaller average node criticality $E[\omega]$ implies a more robust network.

7.3.4. TRAFFIC ASSIGNMENT

We investigate four kinds of traffic assignments: all or nothing, modal-split logit model, user equilibrium and system optimum.

ALL OR NOTHING (AON)

An all-or-nothing (AoN) assignment is commonly applied for traffic assignment in networks. The AoN assignment in this chapter assigns all demand of each OD pair to the route with the lowest route time between the OD pair [154][165], where the link weight is the free-flow travel time $t_{0\ell} = \frac{d_\ell}{v_\ell}$. Two limits of the AoN algorithm are: (1) the capacities c_ℓ of the pathways are assumed to be infinite; and (2) the diversity of route choices for each OD pair is not taken into account. Despite the limits of the AoN assignment, the operational simplicity without extra parameters leads to rapid computation, which allows to roughly understand traffic behaviors under low-loaded situations. The shortest routes are computed by the Bellman-Ford algorithm [166] in our framework.

MODAL-SPLIT ASSIGNMENT (MS)

Modal-split assignment also assumes the infinite link capacity c_ℓ. The demands between origins and destinations are distributed over the network by applying a probabilistic route choice model in modal-split assignment [162]. The freight between the origin and the destination is distributed over several different routes in the route choice set by using a multinomial logit regression. Given the routes set $\mathcal{P}_{sd}$ with the first $|\mathcal{P}_{sd}| = K$ shortest paths from origin s to destination d, the fraction of freight using the k-th route $\mathcal{P}_{sd,k}$ is defined as

$$p_{sd,k} = \frac{\exp\left(-\beta t(\mathcal{P}_{sd,k})\right)}{\sum_{k\in K} \exp\left(-\beta t(\mathcal{P}_{sd,k})\right)} \tag{7.7}$$

where $t(\mathcal{P}_{sd,k})$ is the total time along the route $\mathcal{P}_{sd,k}$ alone and β is the parameter tuning the drivers' preference for the route with a lower travel time. The total flow assigned to link ℓ follows $x_\ell = \sum_{s\in\mathcal{N}} \sum_{d\in\mathcal{N}} \sum_{k\in\{1,2,\ldots,K\}} D_{sd} \cdot p_{sd,k} \cdot \delta(\mathcal{P}_{sd,k}, \ell)$, where $\delta(\mathcal{P}_{sd,k}, \ell)$ takes the value one if link ℓ is belong to the route $\mathcal{P}_{sd,k}$ and zero otherwise.

USER EQUILIBRIUM (UE)

According to Wardrop's first principle [167], drivers in a congested network prefer choosing their route selfishly, following a behavior that is captured by the Nash equilibrium of the underlying non-cooperative game. Assuming that the driver has perfect knowledge of the travel time on a network and able to choose the best route according to Wardrop's first principle [168], the behavioral assumption will lead to a deterministic user equilibrium. Using a potential function $\phi_\ell(x_\ell) = \int_0^{x_\ell} t_\ell(x)dx$, this routing behavior minimizes the sum of the potential functions, which is formulated as a convex optimization problem:

$$\begin{aligned} \underset{x_\ell \forall \ell\in L}{\text{minimize}} \quad & \sum_{\ell\in\mathcal{L}} \int_0^{x_\ell} t_\ell(x)dx \\ \text{subject to} \quad & \sum_{\mathcal{P}} f_{\mathcal{P}}^{sd} = \frac{D^{sd}}{h} \\ & x_\ell = \sum_s \sum_d \sum_{\mathcal{P}} f_{\mathcal{P}}^{sd} \delta^{sd}(\mathcal{P}, \ell) \\ & x_\ell \geq 0, f_p^{sd} \geq 0. \end{aligned} \tag{7.8}$$

where x_ℓ is the total flow on link ℓ, $f_{\mathcal{P}}^{sd}$ is the flow between origin s and destination d on route $\mathcal{P}$, and the indicator $\delta^{sd}(\mathcal{P}, \ell) = 1$ if link ℓ belongs to route $\mathcal{P}$, and $\delta^{sd}(\mathcal{P}, \ell) = 0$ otherwise. The above traffic assignment problem can be solved by the Frank–Wolfe algorithm [169]. The total travel time in the UE assignment is computed using $C_G = \sum_{\ell\in L} x_\ell^* h \cdot t_\ell(x_\ell^*)$ where x_ℓ^* is the solution of problem (7.8).

SYSTEM OPTIMUM (SO)

System optimum assignment follows Wardrop's second principle, where drivers cooperate with each other to minimize the total travel cost [163] of the whole transport system. The flow configuration that results in the optimal total travel time refers to the socially optimal flows obtained by the problem:

$$\begin{aligned} \underset{x_\ell \forall \ell \in L}{\text{minimize}} \quad & C = \sum_{\ell \in L} x_\ell h \cdot t_\ell(x_\ell) \\ \text{subject to} \quad & \text{constraints in (7.8).} \end{aligned} \tag{7.9}$$

which also can be solved by a modified Frank–Wolfe algorithm. The total travel time in the SO assignment is computed by $C_G = \sum_{\ell \in L} x_\ell^* h \cdot t_\ell(x_\ell^*)$ with the solution x_ℓ^* of problem (7.9). Some previous works have discussed the relation between the UE and the SO [170][171], which propose that a trade-off between the UE and the SO agrees more with the real-world traffic behaviors. Beyond this chapter, the sensitivity analysis for the UE and the SO with respect to topological changes may help to compute the flow on each link in a perturbed network more rapidly [172][173].

7.4. CASE STUDY: THE DUTCH CONTAINER FREIGHT TRANSPORT NETWORK

7.4.1. NETWORK MODEL AND CONFIGURATION

UNDERLYING TOPOLOGY

We apply the NWB (*Nationaal Wegenbestand*) database to construct the transport network. The NWB is made by Rijkswaterstaat and defines the section and intersections for all public roads[1], waterways[2] and railways[3] in the Netherlands. For each mode of transport a subset of the network is used. Only the large waterways, the railways that are regularly used for freight transport and the national roads are considered for the network. The terminal nodes are defined in the Rijkswaterstaat documents[4,5,6], while the interdependent crossings are located manually according to Google map. One can refer to our previous work [174] for the details about the infrastructure considered and the network modeling. Figure 7.3 illustrates the transport network of the Netherlands.

The Netherlands is divided into 40 regions (NUTS-3 used by Eurostat) based on BasGoed[7], each of which is represented by a centroid node. Figure 7.4 illustrates the amount of containers transported from and towards all regions in the Netherlands, which determines the OD demand matrix D. We assume that all the freights enter the

---

[1] Rijkswaterstaat, "Nwb-wegen [roads]", 2017. URL: https://www.pdok.nl/nl/ service/wfs-nwb-wegen-nationaal-wegen-bestand.

[2] Rijkswaterstaat, "Nwb-vaarwegen [waterways]", 2017. URL: https://www.pdok.nl/nl/service/wfs-nwb-wegen-nationaal-vaarwegen-bestand.

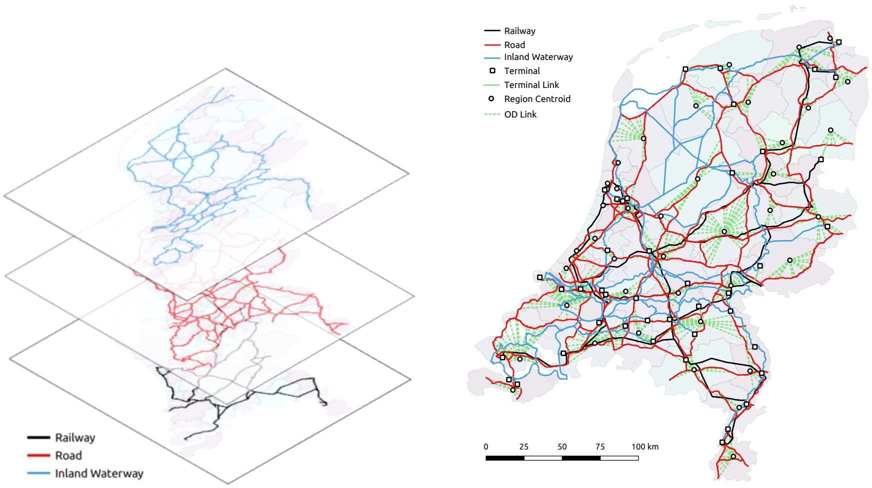
[3] Rijkswaterstaat, "Nwb-spoorwegen [railways]", 2017. URL: https://www.pdok.nl/nl/service/wfs-nwb-spoorwegen-nationaal-wegen-bestand.

[4] ECORYS Nederland BV. "Intermodal links - planner", 2017. URL: https://intermodallinks.com/Planner/.

[5] Port of Rotterdam. "Inlandlinks", 2017. URL:https://www.inlandlinks.eu/nl/terminals/filter.

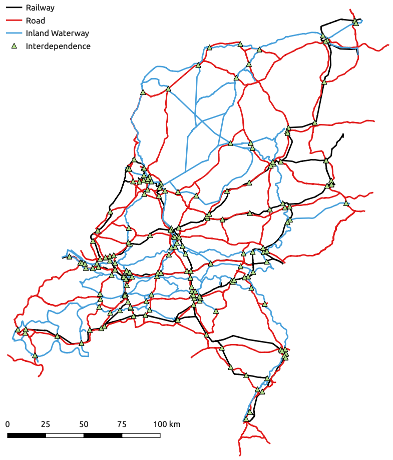
[6] Rijkswaterstaat. "Synchromodaal transport Nederland kaart [synchromodal transport Netherland map]", 2017. Internal document.

[7] Rijkswaterstaat, "Basgoed [model for freight transport]," Internal model, 2017.


Railway
Road
Inland Waterway
Railway
Road
Inland Waterway
Terminal
Terminal Link
Region Centroid
OD Link
0
25
50
75
100 km

(a) Multi-layer network

(b) Main elements


Railway
Road
Inland Waterway
Interdependence
0
25
50
75
100 km

(c) Interdependent nodes

Figure 7.3: Illustration of underlying topology of the freight transport network in the Netherlands. (a) Multi-layered network with three modes of transport. (b) Map of main infrastructure elements. (c) Location of the interdependent nodes.

traffic system in the peak period within one hour, i.e., $h = 1$. The Dutch freight transport network consists of 1457 nodes, 44 terminals, 40 centroids, 1897 main-pathway links, 101 terminal links and 692 OD links.

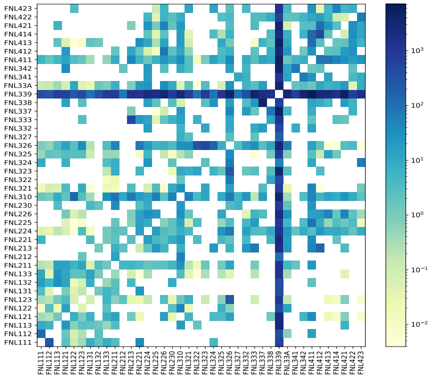


Figure 7.4: Demand matrix D for the amount of containers (in tons per business day) transported in the Netherlands (domestic transport only). The data is from BasGoed.

LINK ATTRIBUTES CONFIGURATION

Different modalities yield different configurations of link attributes in each layer of the network in Figure 7.3a, which define the parameters (i.e., the parameter α, γ, the average speed v_ℓ, the capacity c_ℓ) in the BPR function (7.3) for each link. Table 7.1 presents the link attributes used. The small roads between OD/terminals have a lower average speed than the main roads. Compared with the Road and the Water, the travel time for the Rail depends more on the capacity of the railway, as the maximum number of trains on a railway is predetermined. Therefore, we set a large γ for the Rail. This means the travel time t_ℓ per unit of freight increases little with the freight amount below the capacity, but increases a lot for the amount above the capacity. The information about the average speed and the capacity of each modality is provided in the reports[8],[9],[10].

Modality	α	γ	v_ℓ	c_ℓ
Road (main)	0.15	4	60	2300(vehicle/hour)×2(tons/vehicle)
Road (small)	0.15	4	30	2300(vehicles/hour)×2(tons/vehicle)
Rail	0.15	8	90	2200(tons/train)×2(trains/hour)
Water	0.15	4	15	1200(tons/ship)×5(ships/hour)

Table 7.1: The configuration of the links attributes, including the parameters α and γ, the average speed v_ℓ (in kilometers/hour), and the capacity c_ℓ (in tons/hour).

---

[8] CBS, "Transport of goods in the Netherlands hits new record," 2016. URL: https://www.cbs.nl/en-gb/news/2016/25/ transport-of-goods-in-the-netherlands-hits-new-record.

[9] ProRail, "Network statement 2019," report T20160098-1656408669-827, p. 196, 2017.

[10] Statline, "Traffic intensity; national roads," 2018. URL: https://opendata.cbs.nl/statline/CBS/nl/dataset/82855NED.

7.4.2. ROBUSTNESS ASSESSMENT UNDER RANDOM FAILURES

We next investigate the robustness performance of the network under random failures. Under the scenario of random failure, we remove a fraction of nodes uniformly at random from the network, then compute the increment of the total travel time ΔC_G due to the removals. The ratio of the increment travel time and the original travel time referring to (7.5), i.e., $\eta_G = \frac{\Delta C_G}{C_G}$, is used to measure the robustness of a transport network. Figure 7.5 shows the average indicator $E[\eta_G]$ among all realizations of random node removals under the AoN assignment. The interconnection of multiple modalities can decrease the average normalized increment of the total travel time $E[\eta_G]$, thus improves the robustness for transport services against random failures. Meanwhile, the interdependency effect of the crossing nodes between different modalities could degrade the robustness indicator $E[\eta_G]$ sightly, which is due to the fact that the failure of an interdependent node impacts two modalities simultaneously. Thus, we observe from Figure 7.5 that the interconnection improves the robustness and the interdependency degrades the robustness under random failures.

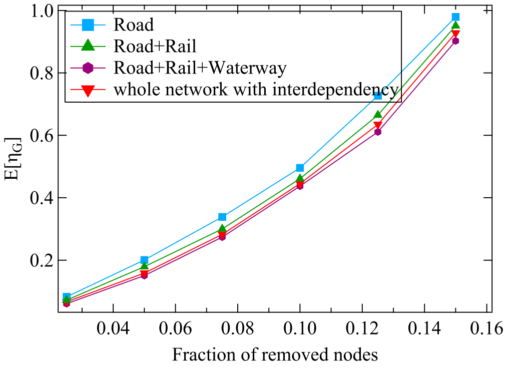
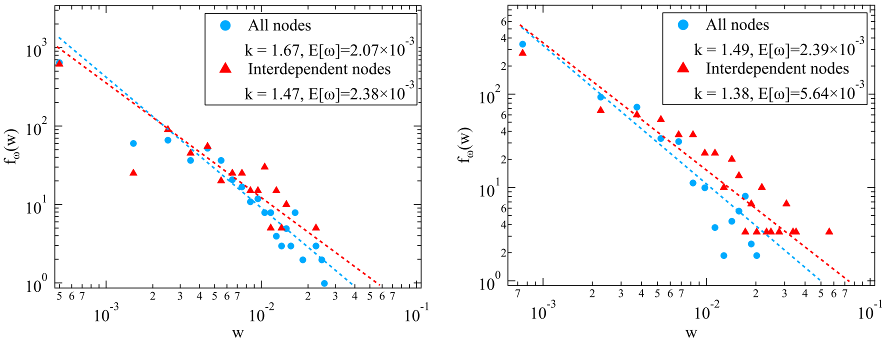


Figure 7.5: The average robustness indicator $E[\eta_G]$ under random failures as a function of the fraction of removed nodes. We compare four cases: (1) the Road, (2) the Road interconnected with the Rail, (3) three interconnected modalities without the interdependent nodes, (4) the whole network. Each data point is based on 100 realizations.

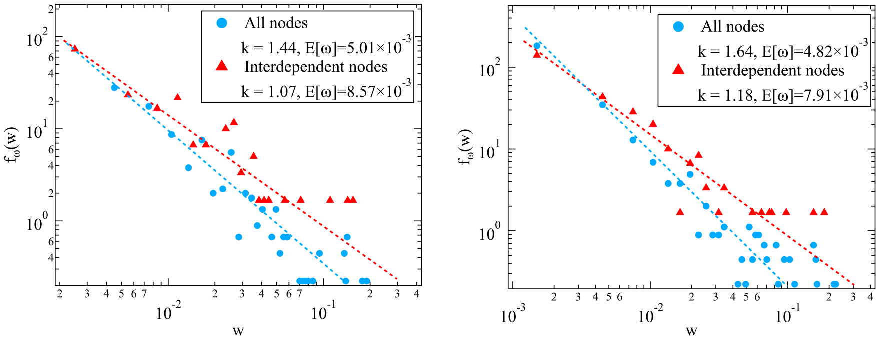
7.4.3. ROBUSTNESS ASSESSMENT VIA NODE CRITICALITY

Next, we investigate the distribution of the node criticality ω for all nodes and the interdependent nodes. Figure 7.6 shows the probability density function (PDF) $f_\omega(w)$ of the node criticality ω as random variables on log-log scale. We fit the distribution of the node criticality ω_i by a power-law PDF that $f_\omega(w) \sim w^{-k}$. We observe from Figure 7.6 that the node criticality resembles a power-law distribution, where different traffic assignment models, i.e., the AoN, the MS2 (i.e. the modal split assignment with $K = 2$ routes), the UE and the SO, have little influence on this scale-free property of robustness. The power-law distribution of the node criticality ω_i implies that the removal of most nodes yields a small increment of the total travel time ΔC_G, while the removal of some critical nodes can increases the travel time C_G significantly. Interestingly, the real-world transport network already tends to be robust against individual failures, which



(a) AoN

(b) MS2



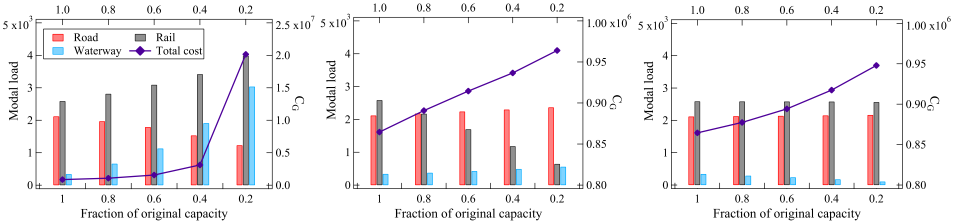
(c) UE

(d) SO

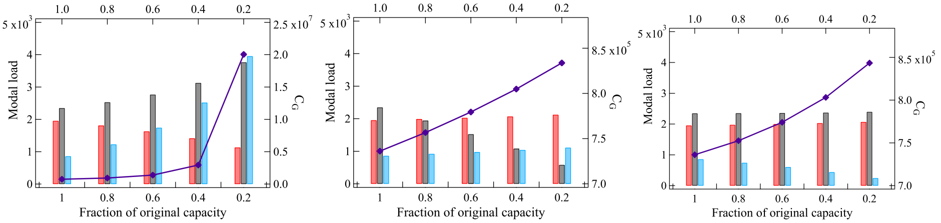
Figure 7.6: Distribution of the node criticality ω for both all nodes and the interdependent nodes. The PDF is fitted by the function $f_\omega(w) \sim w^{-k}$. Both the x-axis and y-axis are on log scales.

may be caused by the topological evolution [175] and traffic optimization during the development of the network.

A larger slope k in the PDF $f_\omega(w) \sim w^{-k}$ implies a better robustness indicated by the node criticality ω of network. Comparing the traffic assignment AoN and the MS2, both with an infinity capacity, the AoN surprisingly presents a better performance against single failures with a larger slope k than the MS2. The AoN employs fewer nodes than the MS2, so that some removed nodes have little influence on the increment of the total travel time ΔC_G. Comparing the UE and the SO, the robustness indicated by the distribution of the node criticality ω under the SO outperforms the UE for disruptions among both all nodes and the interdependent nodes. This implies that a centralized and information-sharing schedule could lead to both a lower total travel time C_G and a better robustness $E[\omega]$. In addition, Figure 7.6 shows that the interdependent nodes usually have a higher node criticality amonge all nodes. Thus the maintenance of the interdependent nodes should be of a higher priority.



(a) Road (UE) (b) Rail (UE) (c) Waterway (UE)



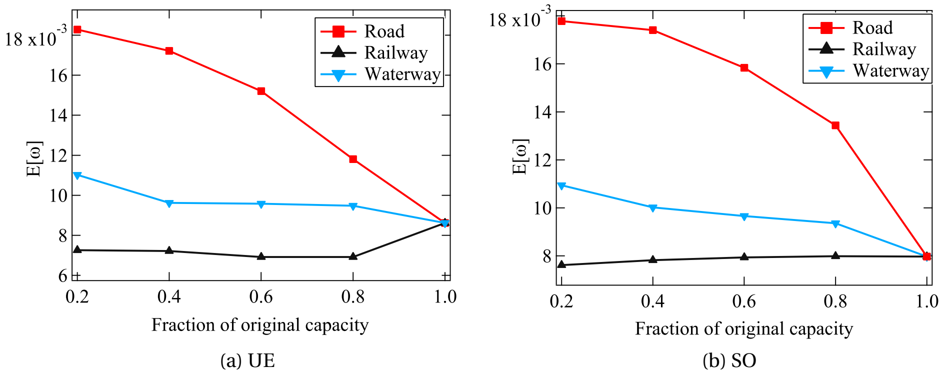
(d) Road (SO) (e) Rail (SO) (f) Waterway (SO)

Figure 7.7: Bar plots show the modal load versus the fraction of the original capacity of a determined modality (the Road, the Rail or the Waterway) and a given traffic assignment (the UE or the SO). The purple lines show the total travel time C_G as a function of the fraction of the original capacity

7.4.4. ROBUSTNESS ASSESSMENT UNDER CAPACITY DEGRADATION

We further investigate the robustness of networks under capacity degradation. First, we define the modal load χ_m as the total flow $\sum_{\ell\in\mathcal{L}_m} x_\ell$ on all the links in this modality divided by the number of links L_m of this modality, i.e. $\chi_m = \frac{\sum_{\ell\in\mathcal{L}_m} x_\ell}{L_m}$, which reflects the usage of each modality. Figure 7.7 shows the modal load χ_m and the total travel time C_G versus the degradation of the capacity of each modality. The degradation of the capacity of a modality decreases the modal load of this modality, while the load shifts to the other two modalities. For the Rail and the Waterway, the total time C_G presents a linear function of the fraction of the original capacity, while the degradation of the Road capacity increases the total travel time C_G sharply. The high sensitivity of the total travel time C_G with the capacity degradation of the Road may be due to the fact that the Road is the dominating modality, and the other two fail to balance the loads if the availability of the Road decreases too much. Figure 7.7 also shows that the SO presents a lower travel time C_G than the UE due to the difference in the usage of Waterway, i.e., the SO has a higher modal load of the Waterway than the UE. This result hints a possible optimization for transportation by making full use of the Waterway.

We consider the single node disruption under capacity degradation of each modality. Figure 7.8 shows the average node criticality $E[\omega]$ as a function of the fraction of the original capacity for each modality, which measures the robustness against single disruption under degradation of capacity. The average node criticality $E[\omega]$ increases with degradation of the capacity of the Road and the Railway and more sensitive to degradation of the Road capacity. By contrast, the average node criticality $E[\omega]$ presents different behaviors for the degradation of the Rail capacity, which more or less presents a Braess's paradox [176], i.e. the robustness degrades for a higher capacity of the Rail.



(a) UE (b) SO

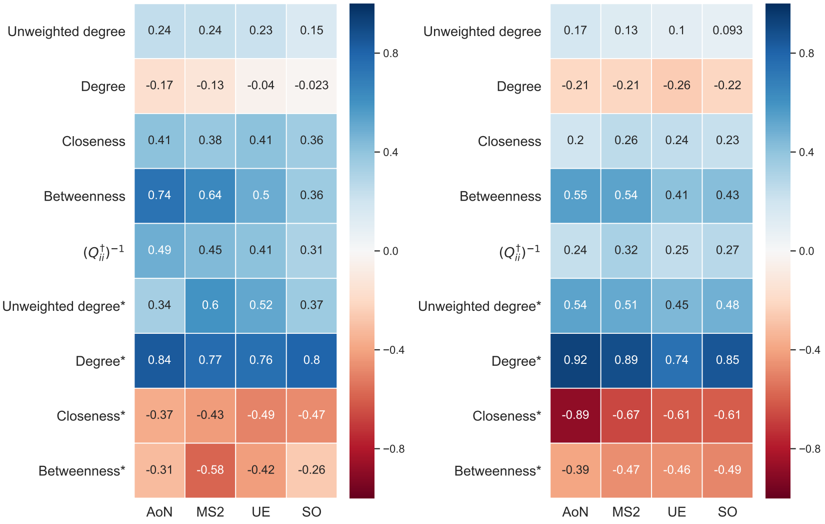
Figure 7.8: The average node criticality $E[\omega]$ as a function of the fraction of the original capacity for each modality under the UE and the SO.

7.4.5. TOPOLOGICAL PROPERTIES OF CRITICAL NODES

We investigate the relation between the nodal topological properties and the node criticality in order to identify the most vulnerable and critical nodes faster [55]. A larger absolute correlation coefficient implies a better nodal metric to identify the critical node. The metric of the unweighted degree is the number of links incident to a node, and the other metrics (i.e. degree, closeness [38], node betweenness [38] and the diagonal element of the pseudo-inverse matrix [64] of the weighted Laplacian matrix $Q^{\dagger}_{ii}$) are computed in the weighted network with the link weight $a_\ell = t_{0\ell} = \frac{d_\ell}{v_\ell}$. The element $Q^{\dagger}_{ii}$ is calculated by the Laplacian matrix $\tilde{Q}$ of the weighted adjacency matrix $\tilde{A}$ with entry $\tilde{a}_\ell = \frac{1}{a_\ell}$ for each link ℓ. Figure 7.9 shows that the betweenness has the highest rank correlation with the node criticality, but the correlation degrades under the MS2, the UE and the SO assignments. Considering that the UE and the SO are more practical in real-world, this degradation implies that the identification of the critical nodes could be difficult if the effect of capacity is taken into account under the UE and the SO.

We define the flow network as the network with the link weight $a_\ell = x_\ell$ and denote the corresponding metrics by (*). The flow network reflects the usage of each link under a specific traffic assignment. The unweighted degree* in the flow network is the number of non-empty links (i.e. $x_\ell \neq 0$) incident to a node. The other metrics are computed in the weighted flow network. The degree* in the flow network is actually the total amount of freight passing this node, which becomes the best indicator of the critical nodes. However, the rank correlation also degrades in the MS2, the UE and the SO, which implies that the critical nodes do not entirely depend on the local traffic flow.

We also investigate the topological properties of critical interdependent nodes. Since the closeness* is calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the flow network [38], the important nodes nearby the links with a high link weight $a_\ell = x_\ell$ usually have a smaller closeness*, which leads to the negative correlations between the node criticality and the closeness*. Figure 7.9 shows that the interdependent nodes present similar behaviour of the topological properties with all nodes, but have a higher correlation with the closeness*. The strong correlation allows us to identify the most critical interdependent nodes by using the degree* and the closeness* in combination.



	AoN	MS2	UE	SO
Unweighted degree	0.24	0.24	0.23	0.15
Degree	-0.17	-0.13	-0.04	-0.023
Closeness	0.41	0.38	0.41	0.36
Betweenness	0.74	0.64	0.5	0.36
$(Q^\dagger_{ii})^{-1}$	0.49	0.45	0.41	0.31
Unweighted degree*	0.34	0.6	0.52	0.37
Degree*	0.84	0.77	0.76	0.8
Closeness*	-0.37	-0.43	-0.49	-0.47
Betweenness*	-0.31	-0.58	-0.42	-0.26

(a) All nodes

	AoN	MS2	UE	SO
Unweighted degree	0.17	0.13	0.1	0.093
Degree	-0.21	-0.21	-0.26	-0.22
Closeness	0.2	0.26	0.24	0.23
Betweenness	0.55	0.54	0.41	0.43
$(Q^\dagger_{ii})^{-1}$	0.24	0.32	0.25	0.27
Unweighted degree*	0.54	0.51	0.45	0.48
Degree*	0.92	0.89	0.74	0.85
Closeness*	-0.89	-0.67	-0.61	-0.61
Betweenness*	-0.39	-0.47	-0.46	-0.49

(b) Interdependent nodes

Figure 7.9: The Spearman's rank correlation coefficients between the node criticality and the topological metrics in the structural network (with $a_\ell = t_{0\ell}$) and the flow network (with $a_\ell = x_\ell$) under different traffic assignments. The coefficients for (a) all nodes and (b) the interdependent nodes are shown in the subgraphs, respectively. We neglect the element $Q^{\dagger *}_{ii}$ in the flow network since the weighted adjacency matrix $\tilde{A}$ is not applicable for the network with zero-weighted links $a_\ell = x_\ell = 0$.

7.5. CHAPTER SUMMARY

Multimodal transport opens a new door for mitigating congestion in road transport and for reducing transportation costs. This chapter addresses the approaches to both the network modeling and the robustness assessment of multimodal transport networks. The consideration of the interdependent property of multimodal networks fills the gap for modeling the disruptions of the crossings. Although the interdependency degrades the robustness of the network slightly, the interconnection of multiple modalities benefits the total travel time and outweighs the negative effect of interdependency on the robustness performance. The robustness is assessed by both element disruptions and capacity degradation. The case study of the Dutch transport network provides several new insights. First, the power-law-like distribution of the node criticality implies a good robustness of the real-world network. The dynamics of self-evolution and overall planning leading to this robust state could be an open question beyond this chapter [177]. Second, we observe that the capacity degradation of the Road could exert a disastrous growth of the total travel time, while shifting more loads to the inland waterways can decrease the total travel time. Lastly, the node criticality is strongly correlated to the amount of freight passing this node. The most critical interdependent node can be identified by the degree and the closeness of nodes in the flow networks. This study can support network planners in tactical and operational decisions for improving the performance of multimodal transport networks.

The proposed framework can be employed to a general transport system, but some details merit further improvements. A better experimental formula for the travel time in terms of the freight flow on links in railway and inland waterways, instead of the BPR function as (7.3), can characterize more features of different modalities. In addition, we employed the total travel time as travel cost for this case study, while different definitions of costs, e.g., CO_2 emissions [154], could exhibit different results.

8

OPTIMIZATION OF CONVERGENCE RATE VIA ALGEBRAIC CONNECTIVITY

The algebraic connectivity of a network characterizes the lower-bound of the exponential convergence rate of consensus processes. This chapter investigates the problem of accelerating the convergence of consensus processes by adding links to the network. Based on a perturbation formula of the algebraic connectivity, we propose a greedy strategy for undirected networks and give a lower bound of its performance through an approximation of submodularity. We further extend our investigation to directed networks, where the second smallest real part among all the eigenvalues of the non-Hermitian Laplacian matrix, i.e., the generalized algebraic connectivity, indicates the expected convergence rate. We propose the metrics to evaluate the impact of an adding subgraph on the generalized algebraic connectivity, and apply a modified greedy strategy to optimize the generalized algebraic connectivity. Numerical results in empirical networks exhibit that our proposed methods outperform some other methods based on traditional topological metrics. Our research also verifies the dramatic differences between the optimization in undirected networks and that in directed networks.

8.1. INTRODUCTION

A Large number of collective dynamic processes, e.g., the Markovian process [38], the network diffusion [178], the fluid flow in tank systems [32], the electric network [64], has a similar objective that all nodes reach an agreement regarding a certain quantity of interest by exchanging the nodal states with their neighboring nodes. These dynamics can be generally described by the consensus model in networks [179, 180]. The convergence rate of a consensus process on a network indicates the speed that each node tends to its final steady state (assuming that the steady state exists), which characterizes the behavior of an autonomous system and can be an important metric of the performance (e.g. robustness or efficiency) for networked control systems.

The behavior of the consensus model can be featured by the eigensystem of the Laplacian matrix of the adjacency matrix, where the algebraic connectivity of the undirected network indicates a lower-bound of the exponential convergence rate [179]. Thus, the problem of convergence acceleration of a process reduces to increasing the algebraic connectivity of the undirected network by topological adjustments. If the variables of interconnections are continues-valued, i.e., the links are weighted, maximizing the algebraic connectivity constrained by a total budget of link weights reduces to a semi-definite programming [181], which can be solved by a subgradient method [182, 183]. For maximizing the algebraic connectivity by adding links in unweighted networks, different heuristic methods based on the topological metrics (e.g., degree, eigenvector centrality, and betweenness) are proposed to approach this NP-hard integer-optimization problem [184–187]. However, the above topological metrics-based approaches cannot theoretically evaluate the performance. It is of practical significance to propose a strategy in undirected networks whose performance can be guaranteed, which is the first issue that we will study in this chapter.

Moreover, several flow processes, e.g., the impedance circuit, the liquid flow, the traffic flow, yields the fact that the interactions between nodes are directed instead of bidirected. The convergence rate to consensus in a directed network depends on both the topology of the network and its initial state vector [188]. The expected convergence rate (ECR) is defined to measure the speed of convergence for random initial state vectors in directed networks [189]. Notwithstanding the importance and generality of the expected convergence rate in directed networks, the method of maximizing the ECR in directed networks by topological adjustments, to the best of our knowledge, has seldom been investigated. Unfortunately, the methods for undirected networks are mostly infeasible for directed networks, which motivates us to design an algorithm to maximize the ECR in directed networks. It is the second issue that we will explore in this chapter.

In this chapter, we investigate the impact of the network perturbation on the convergence rate of dynamics via the algebraic connectivity both in undirected and directed networks, which are confined to unweighted networks. For undirected networks, we present a perturbation formula of the algebraic connectivity for an adding subgraph, which shows the impact of multiple adding links on the algebraic connectivity, as well as demonstrates that the algebraic connectivity resembles an approximately submodular function with respect to the adding subgraphs. Based on the metric for the impact of an individual link, the greedy strategy to maximize the algebraic connectivity by adding links one by one could have a constant performance.

For directed networks, we confine ourselves to the strongly connected network without self-loops to ensure the existence of the non-trivial steady state [180]. We show that the second smallest real part among all the eigenvalues of the Laplacian matrix indicates the lower bound of the expected convergence rate in directed networks, which is defined as the generalized algebraic connectivity [189]. Based on the bounds of the generalized algebraic connectivity perturbation with respect to the adding subgraph, we propose the metrics to measure the impact of an individual link and heuristically apply the greedy strategy to maximize the generalized algebraic connectivity. Further, we compare the performance of the proposed methods with other heuristic methods based on traditional nodal centrality, and discuss the difference between the strategies for increasing the convergence rate between undirected and directed networks.

The main contribution of this chapter can be summarized as:

1. We propose the greedy strategy to maximize the algebraic connectivity μ in undirected networks, which guarantees a constant performance.
2. We propose the heuristic greedy strategy to maximize the generalized algebraic connectivity $\Re(\mu)$ in directed networks.
3. Numerical tests in real-world networks show the superiority of our proposed methods compared with some other heuristics.

The remainder of this chapter is organized as follows. The physical significance of the algebraic connectivity and the definition of the generalized algebraic connectivity are introduced in Section 8.2. Section 8.3 provides a perturbation formula of the algebraic connectivity and propose the greedy algorithm in undirected networks. We extend the optimization problem and propose some heuristic strategies for directed networks in Section 8.4. We evaluate the proposed methods in Section 8.5 and conclude this chapter in Section 8.6.

8.2. ALGEBRAIC CONNECTIVITY FOR CONSENSUS PROCESSES IN NETWORKS

8.2.1. CONSENSUS PROCESSES IN UNDIRECTED NETWORKS

We consider that an undirected and unweighted network $G(\mathcal{N},\mathcal{L})$ consisting of set $\mathcal{N}$ with N nodes and set $\mathcal{L}$ with L links is represented by the adjacency matrix A. The entry of the adjacency matrix $a_{ij} = 1$ if there is a link between node i and node j, and otherwise $a_{ij} = 0$. Let $v_i(t)$ denotes the state of node i at time t, the consensus model follows $\frac{dv_i(t)}{dt} = \sum_{j=1}^{N} a_{ij}(v_i(t) - v_j(t))$, and the vector form for the state vector $\boldsymbol{v}(t) = (v_1(t), v_2(t), \ldots, v_N(t))^T$ follows

$$\frac{d\boldsymbol{v}}{dt} = -(\Delta - A)\boldsymbol{v}(t) = -Q\boldsymbol{v}(t) \tag{8.1}$$

where $\Delta = \text{diag}(d_1, d_2, \ldots, d_N)$ is the diagonal degree matrix (e.g., d_i denotes the degree of node i), and Q is the Laplacian matrix of A.

We assuming that all the eigenvalues λ_k of Q are distinct, i.e., $\lambda_1 < \lambda_2 < \cdots < \lambda_N$, associated with the eigenvector $x_1, x_2, \ldots, x_N$. The state vector $\boldsymbol{v}(t)$ can be written as $\boldsymbol{v}(t) = e^{-Qt}\boldsymbol{v}(0)$ with the initial state vector $\boldsymbol{v}(0)$. The smallest eigenvalue in the symmetric matrix Q is equal to $\lambda_N(Q) = 0$, which corresponds to the steady-state eigenvector $\boldsymbol{\pi} = x_N = \boldsymbol{u}^T$. Thus, the state vector can be written by

$$\boldsymbol{v}(t) = \boldsymbol{\pi} + \sum_{k=2}^{N} e^{-\lambda_k t} x_k x_k^T \boldsymbol{v}(0) \tag{8.2}$$

which implies that the state $v_i(t)$ of each node exponentially converges to the steady state. The relation

$$-\lim_{t\to\infty} \frac{\log|v_i(t) - \pi_i|}{t} \geq \min_{k\leq N-1} \{\lambda_k\} \tag{8.3}$$

implies that the second smallest eigenvalue λ_{N-1} of the Laplacian matrix Q, i.e., the algebraic connectivity, indicates a lower-bound of the expected exponential convergence rate for each nodal state with a random initial state.

8.2.2. CONSENSUS PROCESSES IN DIRECTED NETWORKS

We further consider the dynamic processes in the directed network $G(\mathcal{N}, \mathcal{L})$ with N nodes and L directed links. The entry $a_{ij} = 1$ in the asymmetric adjacency matrix A represents a directed link with source node i and target node j. In the consensus process, the existence of link ℓ_{ji} represents the communication from node i to node j. Thus, the consensus model follows $\frac{dv_i(t)}{dt} = \sum_{j=1}^{N} a_{ji}(v_j(t) - v_i(t))$, and the state vector follows

$$\frac{d\boldsymbol{v}(t)}{dt} = -(\Delta_{in} - A^T)\boldsymbol{v}(t) \tag{8.4}$$

where Δ_{in} is the diagonal nodal in-degree matrix. We define the operator $Q = \Delta_{in} - A^T$ as the generalized Laplacian matrix in directed network[1]. If the network is strongly connected [179], i.e., there exists a path from one node to another for any two nodes, the smallest eigenvalue λ_N of Q is equal to 0, and the system can reach a steady state $\boldsymbol{\pi}$. Assuming that the generalized Laplacian matrix has distinct eigenvalues λ_k and corresponding right- and left- eigenvectors x_k and y_k, the state vector can be generalized as

$$\boldsymbol{v}(t) = \boldsymbol{\pi} + \sum_{k=2}^{N} e^{-\lambda_k t} x_k y_k^T \boldsymbol{v}(0) \tag{8.5}$$

Since the complex eigenvalue for the non-Hermitian matrix Q can be written by $\lambda_k = \Re(\lambda_k) + i\Im(\lambda_k)$ with non-negative real parts $\Re(\lambda_k)$, we rewrite the state vector by $\boldsymbol{v}(t) = \boldsymbol{\pi} + \sum_{k=2}^{N} e^{-\Re(\lambda_k)t - i\Im(\lambda_k)t} x_k y_k^T \boldsymbol{v}(0)$. Denoting $\boldsymbol{e}_i$ the basis vector, the difference between the nodal state $v_i(t)$ and its steady state π_i follows

---

[1] Some researchers [179] prefer applying link ℓ_{ji} to represent communication from node i to node j, then the generalized Laplacian matrix is $Q = \Delta_{out} - A$ with the diagonal nodal out-degree matrix Δ_{out}. Our definition in this chapter has no difference with the previous work for system analysis, but can be more intuitive for topological changing.

$$\begin{aligned} |v_i(t)-\pi_i| &= \boldsymbol{e}_i \left| \sum_{k=2}^{N} e^{-\Re(\lambda_k)t - i\Im(\lambda_k)t} x_k y_k^T \boldsymbol{v}(0) \right| \\ &\le \sum_{k=2}^{N} e^{-\Re(\lambda_k)t} |\boldsymbol{e}_i x_k y_k^T \boldsymbol{v}(0)| \end{aligned} \tag{8.6}$$

which yields that the expected exponential convergence rate has a lower bound

$$-\lim_{t\to\infty} \frac{\log|x_i(t)-\pi_i|}{t} \ge \min_{k\le N-1} \{\Re(\lambda_k)\} \tag{8.7}$$

For brevity, we write the algebraic connectivity as $\mu = \lambda_{N-1}$ in undirected networks, and define the generalized algebraic connectivity $\Re(\mu) = \min_{i<N} \Re(\lambda_i)$ as the second smallest real part among all the eigenvalues of the Laplacian matrix in directed networks.

8.3. ALGEBRAIC CONNECTIVITY IN UNDIRECTED NETWORKS

In this section, we derive an approximation of the algebraic connectivity for the topological perturbation in undirected networks. Further, we propose a greedy method to maximize the algebraic connectivity by adding links or subgraphs.

8.3.1. TOPOLOGICAL PERTURBATION FOR THE ALGEBRAIC CONNECTIVITY

We denote by A the adjacency matrix of the original undirected network $G(\mathcal{N},\mathcal{L})$, and by $A+\Delta A$ the adjacent matrix of the network under perturbation ΔA. The $N\times N$ perturbation matrix ΔA is also symmetric, which is positive-definite for the adding subgraph, and negative-definite for deleting a subgraph. The $N\times N$ Laplacian matrix ΔQ corresponding to ΔA is the perturbation to the original Laplacian matrix Q, which is symmetric and positive-definite for an adding subgraph (negative-definite for a delated subgraph). According to the eigenvalue perturbation theorem [190, 191], the perturbation of the algebraic connectivity follows

$$\Delta\mu = x_{N-1}^T \Delta Q x_{N-1} + \sum_{k\ne N-1}^{N} \frac{(x_k^T \Delta Q x_{N-1})^2}{\lambda_{N-1}-\lambda_k} + \mathcal{O}(\|\Delta Q\|_F^3) \tag{8.8}$$

where x_{N-1} is the Fiedler vector and $\|\cdot\|_F$ is the Frobenius norm. However, estimating the algebraic connectivity perturbation $\Delta\mu$ by (8.8) requires all the eigenvalues and the eigenvectors, whose computational complexity is usually high for large networks, i.e. $\mathcal{O}(N^3)$ by QR decomposition.

We hereby present an approximate of the algebraic connectivity for the topological perturbation ΔA. The Laplacian matrix of a network can be convert into a row-stochastic and non-negative matrix S, by using the transformation $S = I_N - \epsilon Q$, where I_N is the identity matrix and $\epsilon > 0$ is a sufficiently small number. The matrix S is irreducible if the network G is connected when the parameter $\epsilon < \frac{1}{d_{\max}}$ ensures S is non-negative [179]. We denote by γ_1 the left eigenvector of S corresponding to its largest eigenvalue 1. Then, we can construct the matrix $R = S - \frac{\gamma_1\gamma_1^T}{\|\gamma_1\|^2}$, and denote by z the left eigenvector of

R corresponding to the largest eigenvalue $\lambda_1(R)$. The algebraic connectivity μ of Q shifts to the largest eigenvalue of R, and can be computed [192, 193] by

$$\mu(Q) = \frac{1}{\epsilon}(1 - \lambda_2(S)) = \frac{1}{\epsilon}(1 - \lambda_1(R)) \tag{8.9}$$

We suppose that the algebraic connectivity μ becomes $\widetilde{\mu} = \mu + \Delta\mu$ if the Laplacian matrix of a network Q becomes $Q + \Delta Q$. The matrix S after perturbation is $S + \Delta S$ with $\Delta S = -\epsilon\Delta Q$. The eigenvector y_1 corresponding to $\lambda_1(S)$ normalized by $\|y_1\|^2 = 1$ is equal to the eigenvector of Q corresponding to $\lambda_N(Q) = 0$, i.e., $y_1(S) = x_N(Q) = (\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}})^T$. Thus, the eigenvector y_1 does not change for the perturbation, i.e., $y_1(S + \Delta S) = y_1(S)$, which yields the perturbation $\Delta R = \Delta S$.

Denoting z the principle eigenvector of the matrix

$$R = I_N - \epsilon Q - \frac{1}{N}J \tag{8.10}$$

where J is an unit matrix, we further estimate the eigenvector $z + \Delta z$ for a small perturbation $\|\Delta R\|_2 \ll \|R\|_2$ by means of one further iteration of the power method [194] as

$$\begin{aligned} z + \Delta z &\approx \frac{(R+\Delta R)z}{\|(R+\Delta R)z\|_2} = \frac{\lambda_1(R)z + \Delta R z}{\|\lambda_1(R)z + \Delta R z\|_2} = \frac{\lambda_1(R)z + \Delta R z}{\sqrt{\lambda_1^2(R) + 2\lambda_1(R)z^T\Delta R z + z^T\Delta R^T\Delta R z}} \\ &\approx z + \frac{1}{\lambda_1(R)}\Delta R z = z - \frac{\epsilon}{\lambda_1(R)}\Delta Q z \end{aligned} \tag{8.11}$$

We apply a second-order perturbation result as $\Delta\lambda_1(R) \approx \Delta\lambda_1^{(1)}(R) + \Delta\lambda_1^{(2)}(R)$ in discrete calculus. The first term $\lambda_1^{(1)}(R)$ can be obtained by the first-order eigenvalue perturbation [191], i.e., $\Delta\lambda_1^{(1)}(R) = z^T\Delta R z$. The second term $\Delta\lambda_1^{(2)}(R)$ following the Taylor theorem equals

$$\Delta\lambda_1^{(2)}(R) = \frac{1}{2}\Delta^2\lambda_1^{(1)}(R) = \frac{1}{2}\left(\Delta\lambda_1^{(1)}(R+\Delta R) - \Delta\lambda_1^{(1)}(R)\right)$$

where the term $\Delta\lambda_1^{(1)}(R+\Delta R)$ can be approximated by

$$\begin{aligned} \Delta\lambda_1^{(1)}(R+\Delta R) &= \frac{(z+\Delta z)^T\Delta R(z+\Delta z)}{(z+\Delta z)^T(z+\Delta z)} \approx z^T\Delta R z + z^T\Delta R\Delta z + \Delta z^T\Delta R z \\ &= z^T(-\epsilon\Delta Q)z + z^T(-\epsilon\Delta Q)(-\frac{\epsilon}{\lambda_1(R)}\Delta Q z) + (-\frac{\epsilon}{\lambda_1(R)}\Delta Q z)^T(-\epsilon\Delta Q)z \\ &= -\epsilon z^T\Delta Q z + \frac{2}{\lambda_1(R)}z^T(\epsilon\Delta Q)^2 z \end{aligned} \tag{8.12}$$

Therefore, we can obtain that the exact perturbation $\Delta\mu$ of the algebraic connectivity approximates

$$\Delta\mu = -\frac{1}{\epsilon}\Delta\lambda_1(R) \approx z^T\Delta Q z - \frac{\epsilon}{\lambda_1(R)}z^T(\Delta Q)^2 z \tag{8.13}$$

and the algebraic connectivity $\widetilde{\mu}(\Delta Q)$ after a perturbation ΔQ is

$$\widetilde{\mu}(\Delta Q) \approx \mu + z^T \Delta Q z - \frac{\epsilon}{\lambda_1(R)} z^T (\Delta Q)^2 z \tag{8.14}$$

According to the perturbation formula (8.14), we can apply the power iteration method to compute the largest eigenvalue $\lambda_1(R)$ and the principle eigenvector z, and further estimate the algebraic connectivity $\widetilde{\mu}$ for the perturbation ΔQ, which reduces the computational complexity to $\mathcal{O}(N^2)$. Compared with the first-order perturbation $\Delta\mu \approx z^T \Delta Q z = \sum_{i,j \in N} \Delta a_{ij}(z_i - z_j)^2$ proposed in [181], the second term in the derived approximation (8.13) exhibits an additional penalized term. Specifically, the second term

$$z^T (\Delta Q)^2 z = 2 \sum_{i,j \in N} \Delta a_{ij}(z_i - z_j)^2 + 2 \sum_{i,j,k \in N} \Delta a_{ij} \Delta a_{jk} (z_i - z_j)(z_k - z_j) \tag{8.15}$$

implies the effect of multiple links incident to a same node in a perturbation subgraph.

8.3.2. MAXIMIZE THE ALGEBRAIC CONNECTIVITY BY ADDING LINKS

We then consider the problem of maximizing the algebraic connectivity by adding subgraphs or links. Ghosh and Boyd [181] show that maximizing the algebraic connectivity μ in undirected networks by allocating the continues-valued links weights $w_\ell \in [0,1]$ is a convex problem. However, the combinatorial optimization problem that maximizing the algebraic connectivity by adding K unweighted links, i.e., $w_\ell \in \{0,1\}$ is a NP-hard problem, whose computational complexity is $\mathcal{O}\left(\binom{\frac{1}{2}N(N-1)-L}{K} N^3\right)$ by a brute force. An intuitive and straight way is to apply the greedy strategy to adding links for increasing the algebraic connectivity as Algorithm 8.1, which selects the link that increases the algebraic connectivity most in each iteration.

Algorithm 8.1 Original Greedy Method

```
Inputs:
    the undirected network $G$ and the number of links $K$
Initialization:
    Set the solution $S$ as an empty link set
while $|S| \le K$ do
    $\ell_{ij} = \arg\max_{\ell \notin G} \{\mu(G \cup \{\ell_{ij}\})\}$
    $S = S \cup \{\ell_{ij}\}$, $G = G \cup \{\ell_{ij}\}$
end while
Return $S$
```

Definition 8.1 *(Submodularity [195]): A set function f is submodular if for all subsets $A \subset B \subseteq V$ and all elements $s \notin B$, it holds that*

$$f(A \cup \{s\}) - f(A) \ge f(B \cup \{s\}) - f(B) \tag{8.16}$$

Physically, submodularity is a diminishing returns property where adding an element to a smaller set gives a larger gain than adding one to a larger set. Greedy strategy is a suggested method of operational simplicity to approach the near-optima for a monotone submodular maximization [196]. A celebrated results by Nemhauser *et al.* [197] proves that the greedy method provides a good approximation to the optimal solution of the NP-hard optimization problem. Defining the impact of an adding subgraph with the Laplacian matrix ΔQ on the algebraic connectivity as

$$\mathcal{E}(\Delta Q) = z^T \Delta Q z - \theta z^T (\Delta Q)^2 z \tag{8.17}$$

where the constant $\theta := \frac{\epsilon}{\lambda_1(R)}$ depends on the original network, we now investigate the property of the impact function $\mathcal{E}(\Delta Q)$.

In the following analysis, we assume that the impact function $\mathcal{E}(\Delta Q)$ approximates the algebraic connectivity $\Delta\mu$ well, thus the maximzation of $\mathcal{E}(\Delta Q)$ and $\Delta\mu$ are consistent. Further, computing the exact algebraic connectivity $\mu(G \cup \{\ell_{ij}\})$ in Algorithm 8.1 for all candidate links ℓ_{ij} in each iteration still requires high computational cost for large scale networks. According to the impact $\mathcal{E}(\Delta Q)$ of a subgraph, the impact of an individual link ℓ_{ij} can be measured by the metric $\Omega(\ell_{ij}) = |z_i - z_j|$, which can be applied to faster select the best link $\ell_{ij} = \arg\max_{l \notin G}\{|z_i - z_j|\}$ for each iteration in Algorithm 8.2. Coincidentally, the impact (8.15) validates that the second term $z^T(\Delta Q)^2 z$ suggests the same metric $|z_i - z_j|$ for the importance of an individual link. We also mention that the improved perturbation function (8.17) with the second term (8.15) can be applied for more general applications, e.g., adding a subgraph with multiple links in each iteration, or batched greedy optimization [198].

The feasibility of the greedy algorithm 8.2 is an open question. Although the algebraic connectivity is a concave function of continuous-valued link weights, Remark 8.1 shows that the approximated increment of the algebraic connectivity $\mathcal{E}(\Delta Q)$ by adding unweighted subgraphs is unfortunately non-submodular. For conciseness, we notate both the link set and the corresponding Laplacian matrix Q of the graph composed of these links by the same notation Q.

Algorithm 8.2 Heuristic Greedy Method for μ

```
Inputs:
    the undirected network $G$ and the integer $K$
Initialization:
    Set the solution $S$ as an empty link set
while $|S| \le K$ do
    Compute the eigenvector $z$ of $R$ by $Q$
    $\ell_{ij} = \arg\max_{\ell \notin G}\{|z_i - z_j|\}$
    $S = S \cup \{\ell_{ij}\}$, $G = G \cup \{\ell_{ij}\}$
end while
Return $S$
```

Remark 8.1 *Given the perturbation of the Laplacian matrix $Q + \Delta Q$, the function $\mathcal{E}(\Delta Q)$ as* (8.17) *is non-submodular with respect to the set ΔQ of adding subgraphs (links).*

Proof: We define the Laplacian matrix for the perturbation subgraphs I, J, U, V, W following the relation that $I \le J$, $V = I + U$, $W = J + U$ and $Y = J - I$, where the symbol

(+) denotes the union of link sets and (-) indicates the relative complement. We have the relations:

$$\begin{aligned}\mathcal{E}(V)-\mathcal{E}(I) &= z^T(I+U)z-\frac{\epsilon}{\lambda_1(R)}z^T(I+U)^2z-\left(z^TIz-\frac{\epsilon}{\lambda_1(R)}z^TI^2z\right)\\ &= z^TUz-\frac{\epsilon}{\lambda_1(R)}z^TU^2z+\frac{\epsilon}{\lambda_1(R)}z^T(IU+UI)z\\ &= \mathcal{E}(U)-\frac{\epsilon}{\lambda_1(R)}z^T(IU+UI)z \qquad (8.18)\\ \mathcal{E}(W)-\mathcal{E}(J) &= \mathcal{E}(U)-\frac{\epsilon}{\lambda_1(R)}z^T(JU+UJ)z \qquad (8.19)\end{aligned}$$

Thus, we can obtain that

$$(\mathcal{E}(V)-\mathcal{E}(I))-(\mathcal{E}(W)-\mathcal{E}(J))=\frac{\epsilon}{\lambda_1(R)}z^T(YU+UY)z \qquad (8.20)$$

which is always positive only if the matrix U and Y are commuting. Thus, the submodularity of the impact function $\mathcal{E}(\Delta Q)$ is not guaranteed. ☐

We then show that the impact function $\mathcal{E}(\Delta Q)$ is approximately submodular.

Lemma 8.1 *The largest eigenvalue of the Laplacian matrix ΔQ of a graph with K links is upper bounded by $K+1$.*

Proof: The largest Laplacian eigenvalue is upper bounded by the sum of the largest nodal degree and the second largest nodal degree [199]. Thus, the largest Laplacian eigenvalue of a graph with K links has the upper bound $K+1$, which occurs in a star graph. ☐

Lemma 8.2 *Assuming that the adding subgraph with the Laplacian matrix ΔQ consists of K links. The function $\mathcal{E}(\Delta Q)$ is bounded by*

$$\delta z^T\Delta Qz \le \mathcal{E}(\Delta Q) \le z^T\Delta Qz \qquad (8.21)$$

where the constant $\delta := 1-\frac{\epsilon}{\lambda_1(R)}(K+1)$.

Proof: We define the constant $\delta > 0$ such that

$$\delta = \min\frac{z^T\Delta Qz-\frac{\epsilon}{\lambda_1(R)}z^T(\Delta Q)^2z}{z^T\Delta Qz} = 1-\max\frac{\epsilon}{\lambda_1(R)}\frac{z^T(\Delta Q)^2z}{z^T\Delta Qz} \qquad (8.22)$$

The term $z^T(\Delta Q)^2z$ obey the relation

$$|z^T(\Delta Q)^2z| \le ||\Delta Q||\cdot|z^T\Delta Qz| \le \lambda_1(\Delta Q)z^T\Delta Qz \qquad (8.23)$$

Invoking the upper-bound of $\lambda_1(\Delta Q)\le K+1$ in Lemma 8.1, we can obtain that $\delta = 1-\frac{\epsilon}{\lambda_1(R)}(K+1)$ and $\delta z^T\Delta Qz \le \mathcal{E}(\Delta Q)$.

The right inequality can be verified by the fact that $\frac{\epsilon}{\lambda_1(R)}z^T(\Delta Q)^2z \ge 0$ for the positive semi-definite matrix $(\Delta Q)^2$. ☐

Usually, the perturbed subgraph are a small amount of links separately located in a large network, so the eigenvalue $\lambda_1(\Delta Q)$ is relatively small, which practically leads to the constant $0<\delta<1$.

Theorem 8.1 *The greedy algorithm 8.2 for maximizing the impact function $\mathcal{E}(\Delta Q)$ by adding K links (that compose a graph with the Laplacian matrix ΔQ), can guarantee the performance*

$$\mathcal{E}(\Delta Q) \geq \frac{1}{1+\frac{K(1-\delta^2)}{\delta^2}}\left(1-\delta^{2K}\left(1-\frac{1}{K}\right)^K\right)\mathcal{E}_{OPT}(\Delta Q^*) \tag{8.24}$$

if $\frac{1-\delta}{1+\delta} \leq \frac{1}{K}$, where $\mathcal{E}_{OPT}(\Delta Q^*)$ is the optimal impact with the optimal solution set ΔQ^*.

Proof: Lemma 8.2 implies that the impact $\mathcal{E}(\Delta Q)$ is a $\delta-$approximately submodular function [200] which approximates the (sub)modular function $g(\Delta Q) = z^T \Delta Q z$. Invoking the performance bound in Theorem 8 proposed in [201], we can arrive that the performance constant is $\frac{1}{1+\frac{K(1-\delta^2)}{\delta^2}}\left(1-\delta^{2K}\left(1-\frac{1}{K}\right)^K\right)$. ☐

The near optimal solution of the exact increment $\Delta\mu^*$ can be approached by Algorithm 8.2 in some networks. Theorem 8.1 shows that the performance of the greedy algorithm is related to the topological property of the original network indicated by the constant θ, i.e., a smaller $\theta = \frac{\epsilon}{\lambda_1(R)}$ leading a larger δ yields a larger performance constant. Specifically, the impact $\mathcal{E}(\Delta Q)$ approximates to the modular(additive) function $g(\Delta Q)$ for a smaller θ, which implies a better performance for the greedy algorithm. The performance constant also degrades with the increasing number of adding links K. In addition, the computational complexity of Algorithm 8.1 is $\mathcal{O}(K\bar{L}N^2)$, while the computational complexity of Algorithm 8.2 reduces to $\mathcal{O}(K(\bar{L}+N^2))$, where $\bar{L} = \binom{N}{2} - L$ is the number of links in the complement of G.

8.4. GENERALIZED ALGEBRAIC CONNECTIVITY IN DIRECTED NETWORKS

This section extends our investigation to directed networks. We define $\Re(\mu)$ as the second smallest real part among all the eigenvalues of the generalized Laplacian matrix $Q = \Delta_{in} - A^T$ in directed networks without self-loops. The properties of the generalized algebraic connectivity $\Re(\mu)$ in directed networks is more complicated than those in undirected networks. For example, the generalized algebraic connectivity $\Re(\mu)$ is even not monotonic with the number of adding links, and an additional directed link could decrease the generalized algebraic connectivity (see an example in Figure 8.1).

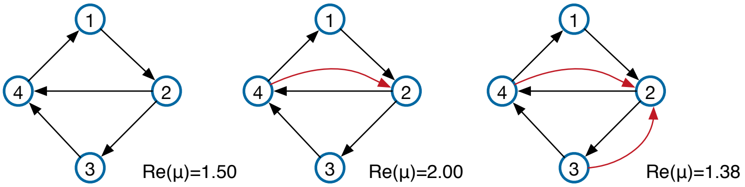


Figure 8.1: Example of the non-monotonicity of the generalized algebraic connectivity $\Re(\mu)$ by adding links.

The algebraic connectivity in a directed network for a real perturbation of Laplacian matrix ΔQ is given by $\Delta\mu = \frac{y^H \Delta Q x}{y^H x} + \mathcal{O}(\|\Delta Q\|_F^2)$, where x and y are right- and left-eigenvectors belonging to μ normalized so that $\|x\|_2 = \|y\|_2 = 1$ and $y^H x = |y^H x|$. For the complex eigenvalue and eigenvectors, we have the perturbation formula

$$\Re(\Delta\mu) = \frac{y_R^T \Delta Q x_R + y_I^T \Delta Q x_I}{y_R^T x_R + y_I^T x_I} + \mathcal{O}(\|\Delta Q\|_F^2) \tag{8.25}$$

where $x = x_R + i x_I$ and $y = y_R + i y_I$. Although the perturbation formula (8.25) provides an estimation of $\Re(\Delta\mu)$, the computational cost of x and y is very high in large networks. Moreover, the change of the real part of a determined eigenvalue $\Re(\Delta\mu)$ could not equal the change of the generalized algebraic connectivity $\Delta\Re(\mu)$.

8.4.1. BOUNDS OF THE GENERALIZED ALGEBRAIC CONNECTIVITY PERTURBATION

Unfortunately, the complex algebraic connectivity μ of a non-Hermitian Laplacian matrix does not obey the power iteration method, which leads that the generalized algebraic connectivity $\Re(\mu)$ perturbation difficult to derive in a similar form with (8.14). Thus, we consider a lower-bound and an upper-bound of the generalized algebraic connectivity $\Delta\Re(\mu)$ under topological perturbation.

Lemma 8.3 *Given the generalized Laplacian matrix $Q = \Delta_{in} - A^T$, the matrix $Q + Q^T$ is positive semi-definite if the network is strongly connected.*

Proof: By the Gershgorin Disk Theorem [202], all the eigenvalues are located in the union of the disks centered at $Q_{jj} + Q_{jj}^T = \sum_{i=1}^N a_{ij} + \sum_{i=1}^N a_{ji}$ with the radius $r_j = \sum_{i=1}^N a_{ij} + \sum_{i=1}^N a_{ji}$ for $j \in \mathcal{N}$. Thus, all the eigenvalues are located in the right plane, which yields that the symmetric matrix $Q + Q^T$ is positive semi-definite. □

Theorem 8.2 *Given a network with the generalized Laplacian matrix Q, the increment of the generalized algebraic connectivity $\Delta\Re(\mu(\Delta Q))$ by adding a subgraph with the generalized Laplacian matrix ΔQ can be lower-bounded by an approximately submodular function, i.e.,*

$$\Delta\Re(\mu(\Delta Q)) \geq \frac{1}{2} z^T \Delta Q^* z - \frac{\epsilon}{4\lambda_1(H)} z^T (\Delta Q^*)^2 z \tag{8.26}$$

where $\Delta Q^* = \Delta Q + \Delta Q^T$, z is the principle eigenvector of $H = \frac{1}{2}(R + R^T)$, R and ϵ are defined in (8.10).

Proof: According to Section 9.3.1, we have the shift relation that $\lambda_i(Q) = \frac{1}{\epsilon}(1 - \lambda_{N-i}(R))$ for the matrix $R = I_N - \epsilon Q - \frac{1}{N} J$. Similarly, we can obtain that

$$\Re(\mu(Q)) = \frac{1}{\epsilon}\left(1 - \max_{i\geq 2} \Re(\lambda_{i-1}(R))\right) \tag{8.27}$$

We then introduce the Bendixson theorem [203] that: let $H = \frac{1}{2}(R + R^T)$ be the Hermitian part of R, we have $\lambda_N(H) \leq \Re(\lambda(R)) \leq \lambda_1(H)$. For the perturbation $Q + \Delta Q$, we have

$\Delta R = -\epsilon \Delta Q$, which yields $\Delta H = -\frac{\epsilon}{2}(\Delta Q + \Delta Q^T) = -\frac{\epsilon}{2}\Delta Q^*$, where ΔQ^* is positive semi-definite due to Lemma 8.3. Invoking the eigenvalue perturbation in (8.14), we arrive at

$$\begin{aligned}\Re(\lambda_{i-1}(R+\Delta R)) &\leq \lambda_1(H+\Delta H)\\ &\approx \lambda_1(H) - \frac{\epsilon}{2} z^T \Delta Q^* z + \frac{\epsilon^2}{4\lambda_1(H)} z^T (\Delta Q^*)^2 z \end{aligned} \tag{8.28}$$

where z is the left eigenvector of the matrix H corresponding to the largest eigenvalue $\lambda_1(H)$ normalized by $\|z\|_2 = 1$. We can obtain the lower bound of the increment of the general algebraic connectivity $\Delta\Re(\mu(\Delta Q))$ follows

$$\begin{aligned}\Delta\Re(\mu(\Delta Q)) &= \Re(\mu(Q+\Delta Q)) - \Re(\mu(Q))\\ &\geq \frac{1}{2} z^T \Delta Q^* z - \frac{\epsilon}{4\lambda_1(H)} z^T (\Delta Q^*)^2 z \end{aligned} \tag{8.29}$$

Since the matrix ΔQ^* is positive definite and Hermitian, the function on the right side of (8.26) is an approximately submodular function (similar with Lemma 8.2), which completes the proof. □.

We also present an upper-bound of the generalized algebraic connectivity perturbation $\Delta\Re(\mu)$ based on Lemma 8.4.

Lemma 8.4 *[189] Defining the matrix for a directed network as* $R = e^{I_N - \alpha Q} - e x_N x_N^T$, *where* x_N *is the right eigenvector of the Laplacian matrix* Q *associated with its zero eigenvalue, and* $0 < \alpha < d_{in}^{\max}$, *the generalized algebraic connectivity holds* $\Re(\mu(Q)) = \frac{1}{\alpha}(1 - \log(\max_{i\in N}|\lambda_i(R)|))$.

According to Lemma 8.4, computing the generalized algebraic connectivity $\Re(\mu)$ can be reduced to computing the maximum absolute eigenvalue of the matrix R by an exponential operator. Then, the generalized algebraic connectivity $\Re(\mu)$ in large scale networks can be computed by a general power iteration method by Krylov subspace [204].

Theorem 8.3 *The increment of the generalized algebraic connectivity* $\Delta\Re(\mu(\Delta Q))$ *by adding a subgraph with the generalized Laplacian matrix* ΔQ *can be upper-bounded by the function*

$$\Delta\Re(\mu(\Delta Q)) \leq \frac{1}{\alpha}\log\frac{\widehat{\lambda}(R)}{\widehat{\lambda}(R) - \kappa(Z)\|\alpha\Delta Q e^{I_N-\alpha Q}\|_p} \tag{8.30}$$

where α *and* R *are defined in Lemma 8.4,* $\kappa(Z)$ *is the condition number of the matrix* Z *composed by the eigenvector of* R, *and* $\widehat{\lambda}(R) = \max_{i\in N}|\lambda_i(R)|$.

Proof: We have the perturbation of the matrix

$$R + \Delta R = e^{I_N - \alpha(Q+\Delta Q)} - e x_N x_N^T \approx (I_N - \alpha\Delta Q)e^{I_N-\alpha Q} - e x_N x_N^T$$

which yields $\Delta R \approx -\alpha\Delta Q e^{I_N - \alpha Q}$ for a small α. According to Lemma 8.4, the problem of maximizing the general algebraic connectivity is equal to minimizing the largest absolute eigenvalue of R. Invoking the Bauer-Fike theorem [205] and the reverse triangle inequality, we can obtain that

$$|\lambda(R)| - |\lambda(R+\Delta R)| \le |\lambda(R) - \lambda(R+\Delta R)| \le \kappa(Z)\|\Delta R\|_p \tag{8.31}$$

where $\kappa(Z) = \|Z\|_p\|Z^{-1}\|_p$ is the condition number [206] of the matrix Z composed by the eigenvector of R.

Denoting the maximum absolute eigenvalue $\max_{i\in N}|\lambda_i(R)|$ by $\widehat{\lambda}(R)$, we have

$$\begin{aligned} \Delta\Re(\mu(\Delta Q)) &= \frac{1}{\alpha}(1-\log\widehat{\lambda}(R+\Delta R)) - \frac{1}{\alpha}(1-\log\widehat{\lambda}(R)) \le \frac{1}{\alpha}\log\frac{\widehat{\lambda}(R)}{\widehat{\lambda}(R)-\kappa(X)\|\Delta R\|_p} \\ &= \frac{1}{\alpha}\log\frac{\widehat{\lambda}(R)}{\widehat{\lambda}(R)-\kappa(X)\|\alpha\Delta Q e^{I_N-\alpha Q}\|_p} \end{aligned} \tag{8.32}$$

which completes the proof. □

8.4.2. MAXIMIZE THE GENERALIZED ALGEBRAIC CONNECTIVITY BY ADDING LINKS

We then consider the problem of maximizing the generalized algebraic connectivity by adding K links in directed networks. Assuming that we apply the greedy strategy to adding the links one by one, a simple and intuitive method to select the link in each iteration follows two steps: 1. obtain the best undirected link regarding the network as undirected; 2. randomly determine the direction of this link. However, the performance of this method is not guaranteed since the effect of the link direction is ignored. The bounds of the perturbation indicate the impact of topological perturbation in two different aspects: the lower bound (8.26) implies that the increment of the generalized algebraic connectivity is related to the connections between nodes with different eigenvector centrality, while the exponential norm of the perturbation $\|\Delta R\|$ influences the upper bound (8.30).

Further, Theorem 8.2 demonstrates that the maximal impact of an individual additional link on the lower bound of the increment of the general algebraic connectivity $\Delta\Re(\mu(\Delta Q))$ can be measured by

$$\underline{\Omega}(\ell_{ij}) = \frac{1}{2}z_j(z_j - z_i) - \frac{\epsilon}{4\lambda_1(H)}\left((2z_j - z_i)^2 + z_j^2\right) \tag{8.33}$$

which is different from the impact $\Omega(\ell_{ij}) = |z_i - z_j|$ in undirected networks. The dominating first term $\frac{1}{2}z_j(z_j - z_i)$ in (8.33) implies that the best link ℓ_{ij} tends to be located in two different communities to maximize the difference $z_j - z_i$. Meanwhile, for the directed links between the same two nodes, i.e., ℓ_{ij} and ℓ_{ji}, the target of the additional link tends to be the node with a higher value z_j to increase the generalized algebraic connectivity. Theorem 8.3 demonstrates that the impact of the subgraph depends on $\|\Delta R\|_p$. Defining the matrix $W = e^{I_N-\epsilon Q} \approx 2I_N - \epsilon Q + \frac{1}{2}(I_N - \epsilon Q)^2$ and applying norm-1 for computational simplicity, we can obtain another metric to measure the importance of an individual link ℓ_{ij} on the generalized algebraic connectivity, i.e.,

$$\overline{\Omega}(\ell_{ij}) = \|\Delta Q e^{I_N-\epsilon Q}\|_1 = \max_{k\in N}|W_{ik} - W_{jk}| \tag{8.34}$$

We assume that the impact of an adding link on the bounds of $\Delta\Re(\mu)$ has similar behaviors with the exact generalized algebraic connectivity. By adding the link with the largest impact in each iteration, we heuristically propose the greedy method as Algorithm 8.3 to approach the near-optimal solution.

Algorithm 8.3 Heuristic Greedy Method for $\Re(\mu)$

```
Inputs:
    the undirected network $G$ and the integer $k$
Initialization:
    Set the solution $S$ as an empty link set
while $|S| \le k$ do
    Compute matrix $R$ by the Laplacian matrix $Q$
    $\ell_{ij} = \arg\max_{l \notin G} \underline{\Omega}(\ell_{ij})$ or $\ell_{ij} = \arg\max_{l \notin G} \overline{\Omega}(\ell_{ij})$
    $S = S \cup \{\ell_{ij}\}$, $G = G \cup \{\ell_{ij}\}$
end while
Return $S$
```

8.5. NUMERICAL EVALUATIONS

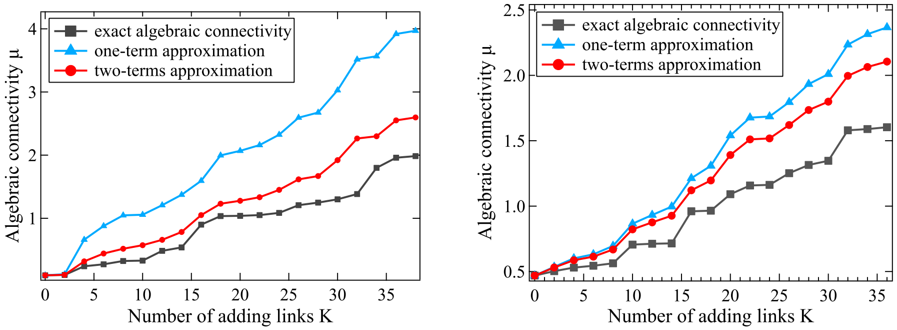
In this section, we present some numerical tests to evaluate the performance of the perturbation formula and our proposed methods. The topological properties of the investigated networks are summarized in Table 8.1. We extract the giant component from the undirected networks and extract the largest strongly connected component from the directed networks.

	N	L	λ_1	$\Re(\mu)$	Type
Karate [94]	34	78	6.73	0.469	undirected
Les Misérables [71]	77	254	12.00	0.205	undirected
NetScience [73]	379	914	10.38	0.015	undirected
Illinois friendship [207]	67	359	5.780	0.115	directed
Berlin traffic [208]	216	514	3.35	0.022	directed
Neural network [209]	239	1912	9.15	0.364	directed

Table 8.1: The topological properties of the giant component or the largest strongly connected component of several experimental networks

8.5.1. UNDIRECTED NETWORKS

We first evaluate the perturbation formula (8.14) of the algebraic connectivity by continuously adding links in undirected networks. In each step, we randomly select one node and randomly adding two new links incident to this node. Figure 8.2 shows the exact algebraic connectivity as a function of the number of the adding links K, which is compared with the perturbation formula with one term, i.e., $\widetilde{\mu} \approx \mu + z^T \Delta Q z$ and the perturbation formula (8.14) with two terms $\widetilde{\mu} \approx \mu + \Delta\mu$. We observe that the algebraic connectivity estimated by the perturbation formula (8.14) usually provides an upper-bound of the exact algebraic connectivity μ, but deviates more from the exact algebraic connectivity μ with the increasing number of adding links, which implies that the performance of the perturbation formula could degrade with the increasing size of the adding subgraphs. The proposed perturbation formula (8.14) can provide a better estimation for the exact algebraic connectivity μ, especially for the spare network with a small maximum degree (e.g., circle networks). The proposed approximation (8.14) exhibits a similar behavior with the perturbation with one term in a dense network since the small $\epsilon < \frac{1}{d_{\max}}$ of (8.14) reduces the impact of the second term.



(a) Circle with 20 nodes

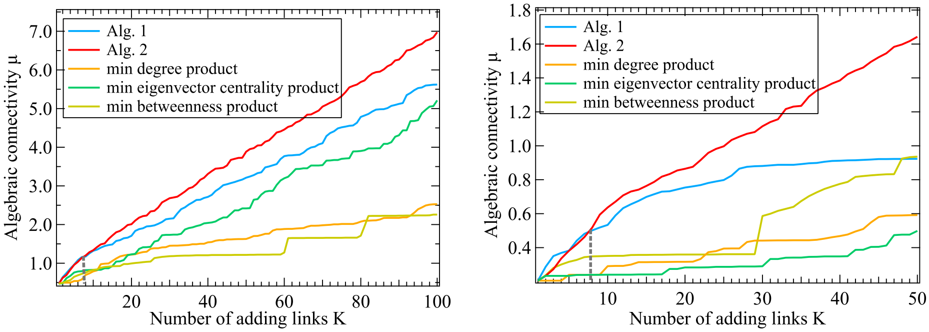
(b) Karate

Figure 8.2: The functions of (i) the exact algebraic connectivity, (ii) the approximation with one term $\widetilde{\mu} \approx \mu + z^T \Delta Q z$ and (iii) the approximation (8.14) with two terms with respected to the number of adding links K.

We then compare the performance of different strategies to maximize the algebraic connectivity, which includes the strategies for selecting a link in each iteration:

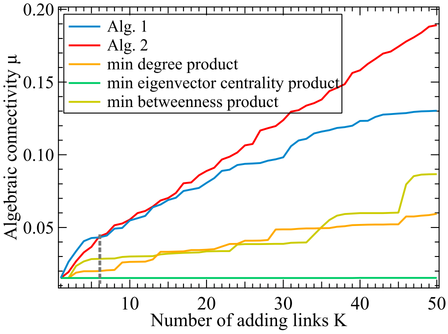
(1) Alg. 1: selecting the link ℓ_{ij} to maximize μ greedy;
(2) Alg. 2: selecting the link ℓ_{ij} with $\max\{|z_i - z_j|\}$;
(3) Selecting the link ℓ_{ij} between the nodes with the smallest degree product $d_i d_j$;
(4) Selecting the link ℓ_{ij} between the nodes with the smallest eigenvector centrality product;
(5) Selecting the link ℓ_{ij} between the nodes with the smallest node betweenness product.

Figure 8.3 shows the exact algebraic connectivity μ as a function of the number of adding links K via different greedy strategies in three empirical undirected networks. Figure 8.3 shows that the algebraic connectivity μ via Algorithm 8.1 exhibits a concave-like function with respect to the number of adding links K, and has the best performance for a small number of adding links (e.g., $K \approx 5$). The proposed heuristic method Algorithm 8.2 approaches the performance of Algorithm 8.1 best, and even outperforms Algorithm 8.1 when the number of adding links K is large. We also observe that the other heuristic greedy strategies based on the topological metrics (e.g., degree, eigenvector centrality and betweenness) usually cannot guarantee the performance for maximizing the algebraic connectivity.



(a) Karate

(b) Les Misérables



(c) Netscience

Figure 8.3: The exact algebraic connectivity μ as a function of the number of adding links K via different greedy strategies in three empirical undirected networks, e.g, Karate, Les Misérables and Netscience.

8.5.2. DIRECTED NETWORKS

We further evaluate the performance of the heuristic metrics proposed in Section 9.4.3 in directed networks. We consider the following metrics to select the individual link in each iteration in the greedy strategy:

(1) Maximize $\Re(\mu)$: the link that maximizes the generalized algebraic connectivity $\Re(\mu)$;
(2) Maximize $\underline{\Omega}(\ell_{ij})$: the link with the maximum $\underline{\Omega}(\ell_{ij})$;
(3) Maximize $\underline{\Omega}(\ell_{ij})$ with inverse direction: the inverse directed link of the link via (2);
(4) Maximize $\overline{\Omega}(\ell_{ij})$: the link with the maximum $\overline{\Omega}(\ell_{ij})$;
(5) Undirected handling: first select the undirected link with the maximum $\Omega(\ell_{ij})$ by regarding the network is undirected, and then randomly determine the link direction.

Figure 8.4 shows that the method “maximize $\Re(\mu)$" presents the best performance for a small number of adding links K although the submodularity of the function $\Delta\Re(\mu)$ in directed networks is not guaranteed. However, the generalized algebraic connectivity

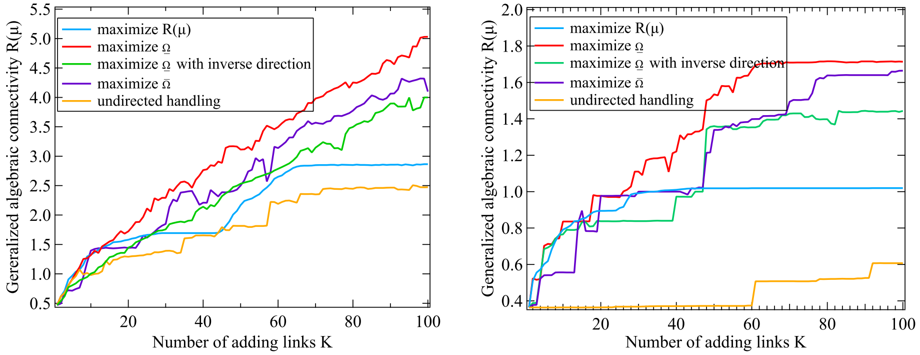
$\Re(\mu)$ via the method "maximize $\Re(\mu)$" tends to increase monotonically but very slowly when the number of adding links is large, which implies that the original greedy method of "maximize $\Re(\mu)$" could lead to the local optima in directed networks. The method " maximize $\underline{\Omega}(\ell_{ij})$ " is the best heuristic method to approach the method "maximize $\Re(\mu)$" for a small fraction of adding links, and performs better for a large fraction of adding links.

We observe that the generalized algebraic connectivity $\Delta\Re(\mu)$ via the method "maximize $\underline{\Omega}(\ell_{ij})$" as a function of the number of adding links K does not smoothly monotonically increase, which may instead help the solution to avoid local optimum possibly due to the inheritance of stochastic optimization (e.g., the principle of "simulated annealing" algorithm and extremal optimization [210]). Also, the method "maximize $\overline{\Omega}(\ell_{ij})$ " can also avoid the local optimum and leads to a good solution, while the performance could degrade much in very sparse networks (e.g., Berlin traffic network with the average degree $E[D] \approx 2.29$).

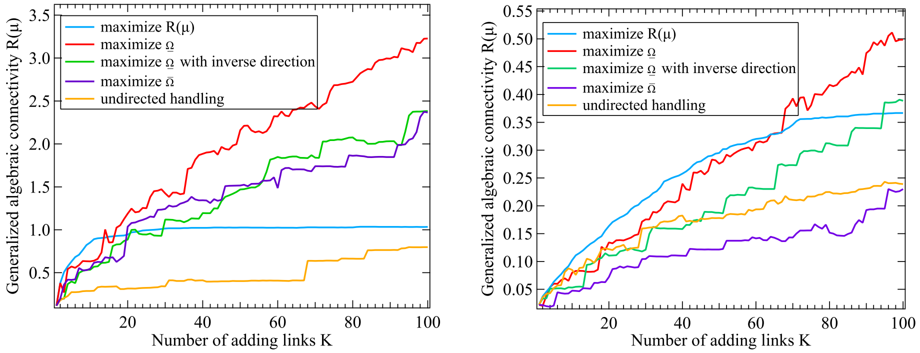
Figure 8.4 also shows the difference of performance between the method "maximize $\underline{\Omega}(\ell_{ij})$ " and the same link with inverse direction, which demonstrates that the link direction is influential to the generalized algebraic connectivity $\Re(\mu)$. Furthermore, the unsatisfied performance of the method "undirected handling" hints the difference of the optimization between directed and undirected networks, as well as the fact that the strategy for undirected networks may be infeasible for directed networks.

8.6. CHAPTER SUMMARY

The (generalized) algebraic connectivity indicates the lower-bound of the exponential convergence rate for consensus processes on networks. This chapter investigated the strategy to maximize the (generalized) algebraic connectivity by adding links in both directed and undirected networks. The approximate submodularity of the derived perturbation of the algebraic connectivity validates that the greedy strategy could guarantee a constant performance in some undirected networks. For directed networks, we proposed the heuristic metrics for selecting the best directed links in the greedy algorithm based on the bounds of the generalized algebraic connectivity perturbation. Numerical tests show that the original greedy strategy based on the exact algebraic connectivity performs best for a small number of adding links, while the greedy strategy based on the proposed metrics could outperform others when the number of adding links is large as well as requires less computational cost. This chapter conveys the insight into the difference on the optimization in directed and undirected networks. Beyond this chapter, the similarities and differences of the behavior of dynamic processes in directed and undirected network merit further study, which is also of practical significance to the community detection for dynamics [211].



(a) Karate (b) Neural network



(c) Illinois friendship (d) Berlin traffic network

Figure 8.4: The exact generalized algebraic connectivity $\Re(\mu)$ as a function of the number of adding links K via different greedy strategies in four empirical undirected networks, e.g., bi-directed Karate, Neural network, Illinois friendship and Berlin traffic network.

9

CONCLUSION

“Was Vernünftig ist, das ist Wirklich; und was Wirklich ist, das ist Vernünftig.”

— Grundlinien der Philosophie des Rechts 1820

9.1. MAIN CONTRIBUTIONS

This thesis contributes the original frameworks, the new insights and the optimization approaches on the performances of complex networks, encompassing the efficiency of spread and the robustness of transport. We are devoted to a better understanding on how the interplay of underlying topology and overlying dynamics affects the behavior of processes and services in networks via both theoretical analysis and case study. The network optimization is a multi-dimensional and multi-objective problem. Structural properties of networks, time-dependent behavior of dynamics, optimization strategy, performance assessment, all need to be taken into account for designing a better network. Revisiting the research questions mentioned in Chapter 1, the main contributions of each chapter are as follows:

We explore time-dependent behaviors of epidemics by the defined spreading time in Chapter 2. The spreading time indicates the time when a spreading process reaches the metastable state and features the velocity of a spread during the outbreak period. We observe that the spreading time in SIS epidemics resembles a lognormal-like distribution with different deep tails, which is exhibited both in the Markovian and the non-Markovian infection process. The deep tail in the distribution of the spreading time implies that the transient increasing period of an epidemic process could be very long. The average spreading time is not necessarily monotonous with the effective infection rate but exhibits a maximum, which means that a higher effective infection rate may not lead to a shorter average spreading time.

Chapter 3 addresses the research question of identifying the fastest initial spreader with the shortest average spreading time for an efficient spreading. With the increasing effective infection rate, the fastest spreader changes from the node with the largest degree to the node with the shortest flooding time. This fact implies that the fastest spreader is coupled to not only the underlying graph but also the dynamic process. Further, we propose the spreading efficiency as a metric to identify the fastest spreaders by considering the expansion and the largest eigenvalue of the subgraph around the spreader.

Chapter 4 introduces the induced SIS spreading in networks, which aims to steer the viruses to the target nodes as much as possible by allocating spreading resources on nodes optimally. We provide two frameworks for the optimal induced spreading: the static optimization and the dynamic optimization. In the static optimization, the optimal infection rate increment of each node is highly related to the degree as well as its average hops to the targets. In the dynamic optimization, the time-dependent optimal infection rate increment exhibits two periods for a large cost budget: steering the viruses from the initial spreader to the target, and maintaining the infection probability of the target by its neighbors. The advantage of the dynamic optimization over the static optimization under a small budget of resources is verified. Also, the differential evolution algorithm presents promising feasibility in high-dimensional network optimizations. We also illustrate that the cost scaling with the fraction of targets has different behaviors compared to that of the targeted controlling in linear systems, because the cost for increasing the infection rate of one node usually benefits the infection probabilities of multiple targets.

The NIMFA steady-state prevalence can be represented by a Taylor series in terms of the effective infection rate at the NIMFA epidemic threshold. Chapter 5 shows that the network topology alters the radius of convergence of the prevalence expansion, which is infinite in regular graphs and becomes finite in irregular graphs. The average radius of convergence increases with the density (the average degree) in random graphs, e.g., ER random graphs and scale-free graphs. The radius of convergence is also coupled to the eigenvalues of the adjacency matrix. Especially, a smaller spectral gap usually decreases the radius of convergence in sparse networks and clustered networks.

Chapter 6 addresses the problem of how to access network recoverability. We propose a topological approach for evaluating the network recoverability in two scenarios, link-based Scenario A and energy-based Scenario B, which extends the application of the framework [17] of network robustness. We assess the recoverability of 10 real communication networks for two different path-based robustness metrics, i.e. the network efficiency and the effective graph resistance. Compared with other metric-based recovery strategies, the greedy recovery strategy exhibits superior performance for the investigated robustness metrics and thus improves the network recoverability more effectively.

Chapter 7 addresses the approaches on both the network modeling and the robustness assessment of multimodal transport networks. The concepts of interconnection and interdependency in network science are introduced to abstract multimodal transport systems as network models better. The consideration of the interdependent effect fills the gap for modeling the disruptions of the crossings. The framework allows us to evaluate the network robustness by simulating link removals and calculating travel cost increments. The results in the case study on the Dutch transport network present the power-law-like distribution of the node criticality, which implies the scale-free robustness against disruptions in the real-world network. The node criticality is strongly correlated to the amount of freight passing this node. The most critical interdependent nodes can be identified by the degree and the closeness in the overlying flow networks.

Chapter 8 demonstrates that the (generalized) algebraic connectivity determines the convergence rate of consensus processes in undirected (or directed) networks. In order to improve the convergence rate of consensus processes, we propose the strategy to maximize the (generalized) algebraic connectivity by adding links in both directed and undirected networks. The approximate submodularity of the derived perturbation of the algebraic connectivity validates that the greedy strategy could guarantee a constant performance in some undirected networks. In directed networks, the greedy algorithm, which selects the directed link based on the proposed heuristic metric in each step, exhibits the advantage over other topological metric-based strategies.

9.2. DIRECTIONS FOR FUTURE WORK

The research questions of this thesis and the above insights obtained from the results open doors to a few future research directions.

Despite the fact that epidemic models have been studied in recent decades, the time-dependent behavior of epidemics in networks still lacks extensive investigations. Chapter 2 demonstrates that the spreading time resembles a lognomal-like distribution, which is due to the interplay between the exponential reproduction of viruses and the Gaussian-distributed viral density in the metastable state. The lognormal distribution commonly appears in the temporal properties of universal human behavior and social activities, such as interactivity time on online social networking [212], duration of strikes [213] and Sartwell's incubation period [214]. Some previous research [215][216] has already tried to explain the incubation time by some dynamic processes. If we interpret the viral density in the metastable state as the threshold of symptom, the similarity between the spreading time and the incubation period may imply the behaviors of viruses spread in individuals. We expect that the analysis and optimization approaches can be applied further if the spreading behaviors in individuals are verified.

Chapter 3 and Chapter 8 provide some insights into optimization approaches for dynamics in networks. Due to the large size of complex networks, the network optimization problems are usually extremely high dimensional. Greedy strategy, with a low computing cost and an operational simplicity, could guarantee a constant performance based on the submodularity theory. Summars *et al.* [217] summarized some submodular indicators for network controllability. An exhaustive investigation on the (approximated) submodularity of the network properties is needed in order to verify the feasibility of greedy strategy in more general cases. Besides the greedy algorithm, a better heuristic algorithm for optimization in large networks is still an open question, though the differential evolution algorithm seems promising, as shown in Chapter 3. Further, considering time-dependent processes on networks, a faster dynamic optimization for networks is also necessary, which can benefit the adaptive dynamic programming for network control [218], as well as the reinforcement learning in neural networks [219].

Nonlinearity and non-Markov property characterize more generic individual behaviors and collective dynamics in the real world, which meanwhile complicates the analysis of processes on networks. Topological heterogeneity, interplaying with the nonlinearity and the non-Markov property, has been shown to induce novel behavior in non-equilibrium dynamics on networks, e.g., the Griffiths phases in contact process [220] and the localization in fracture processes [221]. In addition, we usually assume the constant infection rates and curing rates in epidemic models. However, the practical infection could be time-dependent, state-dependent (e.g. adaptive spreading models [222]), or even history-dependent (with memory effect). In these cases, the behaviors of these models could be dramatically different. Thus, we encourage more efforts on network performances with nonlinearity and non-Markov property in future work.

REFERENCES

[1] M. G. Bell and Y. Iida, *Transportation network analysis* (1997).

[2] I. Simonsen, L. Buzna, K. Peters, S. Bornholdt, and D. Helbing, *Transient dynamics increasing network vulnerability to cascading failures,* Physical Review Letters **100**, 218701 (2008).

[3] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, *Breakdown of the internet under intentional attack,* Physical Review Letters **86**, 3682 (2001).

[4] M. D. Humphries, K. Gurney, and T. J. Prescott, *The brainstem reticular formation is a small-world, not scale-free, network,* Proceedings of the Royal Society B: Biological Sciences **273**, 503 (2005).

[5] M. E. Newman, *The structure of scientific collaboration networks,* Proceedings of the National Academy of Sciences **98**, 404 (2001).

[6] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, *Measurement and analysis of online social networks,* in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (ACM, 2007) pp. 29–42.

[7] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, *Functional connectivity in the resting brain: a network analysis of the default mode hypothesis,* Proceedings of the National Academy of Sciences **100**, 253 (2003).

[8] P. Erdős and A. Rényi, *On random graphs,* Publicationes Mathematicae **6**, 290 (1959).

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

[10] A.-L. Barabási and E. Bonabeau, *Scale-free networks,* Scientific American **288**, 60 (2003).

[11] W. Yu, G. Chen, and J. Lü, *On pinning synchronization of complex dynamical networks,* Automatica **45**, 429 (2009).

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

[13] S. Shao, X. Huang, H. E. Stanley, and S. Havlin, *Percolation of localized attack on complex networks,* New Journal of Physics **17**, 023049 (2015).

[14] J. Villermaux, *Mixing in chemical reactors,* (ACS Publications, 1983).

[15] X. F. Wang and G. Chen, *Pinning control of scale-free dynamical networks,* Physica A: Statistical Mechanics and its Applications **310**, 521 (2002).

[16] D. Kempe, J. Kleinberg, and É. Tardos, *Maximizing the spread of influence through a social network,* in *9th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2003) pp. 137–146.

[17] P. Van Mieghem, C. Doerr, H. Wang, J. M. Hernandez, D. Hutchison, M. Karaliopoulos, and R. Kooij, *A framework for computing topological network robustness,* Delft University of Technology, Report20101218 (2010).

[18] R. Albert, H. Jeong, and A.-L. Barabási, *Error and attack tolerance of complex networks,* Nature **406**, 378 (2000).

[19] S. Broadbent and J. Hammersley, *Mathematical proceedings of the cambridge philosophical society,* , 629 (1957).

[20] R. Cohen, D. Ben-Avraham, and S. Havlin, *Percolation critical exponents in scale-free networks,* Physical Review E **66**, 036113 (2002).

[21] A. Majdandzic, B. Podobnik, S. V. Buldyrev, D. Y. Kenett, S. Havlin, and H. E. Stanley, *Spontaneous recovery in dynamical networks,* Nature Physics **10**, 34 (2014).

[22] J. Gao, B. Barzel, and A.-L. Barabási, *Universal resilience patterns in complex networks,* Nature **530**, 307 (2016).

[23] M. E. Newman, A.-L. E. Barabási, and D. J. Watts, *The structure and dynamics of networks.* (Princeton University Press, 2006).

[24] D. Bernoulli, *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir,* Histoire de l'Acad., Roy. Sci.(Paris) avec Mem , 1 (1760).

[25] R. Pastor-Satorras and A. Vespignani, *Epidemic dynamics and endemic states in complex networks,* Physical Review E **63**, 066117 (2001).

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

[27] C. Li, R. van de Bovenkamp, and P. Van Mieghem, *The SIS meanfield N-intertwined and Pastor-Satorras & Vespignani approximation: a comparison,* Physical Review E **86**, 026116 (2012).

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

[29] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas, *Optimal resource allocation for network protection against spreading processes,* IEEE Transactions on Control of Network Systems **1**, 99 (2014).

[30] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, *Identification of influential spreaders in complex networks,* Nature Physics **6**, 888 (2010).

[31] W. M. Deen, *Analysis of transport phenomena*, Vol. 2 (Oxford university press New York, 1998).

[32] W. L. Luyben, *Process modeling, simulation and control for chemical engineers* (McGraw-Hill Higher Education, 1989).

[33] S. Rai and D. P. Agrawal, *Distributed computing network reliability* (Los Alamitos, CA (USA); IEEE Computer Society Press, 1990).

[34] A. Clark, L. Bushnell, and R. Poovendran, *A supermodular optimization framework for leader selection under link noise in linear multi-agent systems,* IEEE Transactions on Automatic Control **59**, 283 (2013).

[35] Z. He and P. Van Mieghem, *The spreading time in SIS epidemics on networks,* Physica A: Statistical Mechanics and its Applications **494**, 317 (2018).

[36] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, *Emotions as infectious diseases in a large social network: the SISa model,* Proceedings of the Royal Society of London B: Biological Sciences **277**, 3827 (2010).

[37] L. Buzna, K. Peters, and D. Helbing, *Modelling the dynamics of disaster spreading in networks,* Physica A: Statistical Mechanics and its Applications **363**, 132 (2006).

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

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

[40] P. Van Mieghem and R. Van de Bovenkamp, *Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks,* Physical Review Letters **110**, 108701 (2013).

[41] R. Van de Bovenkamp and P. Van Mieghem, *Survival time of the susceptible-infected-susceptible infection process on a graph,* Physical Review E **92**, 032806 (2015).

[42] A. Ganesh, L. Massoulié, and D. 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,* Vol. 2 (IEEE, 2005) pp. 1455–1466.

[43] P. Van Mieghem, *Approximate formula and bounds for the time-varying susceptible-infected-susceptible prevalence in networks,* Physical Review E **93**, 052312 (2016).

[44] M. Draief, *Epidemic processes on complex networks,* Physica A: Statistical Mechanics and its Applications **363**, 120 (2006).

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

[46] J. R. Artalejo, *On the time to extinction from quasi-stationarity: A unified approach,* Physica A: Statistical Mechanics and its Applications **391**, 4483 (2012).

[47] R. Van de Bovenkamp and P. Van Mieghem, *Time to metastable state in SIS epidemics on graphs,* in *Tenth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)* (IEEE, 2014) pp. 347–354.

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

[49] E. Cator and P. Van Mieghem, *Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis,* Physical Review E **87**, 012811 (2013).

[50] E. Cator, R. Van de Bovenkamp, and P. Van Mieghem, *Susceptible-infected-susceptible epidemics on networks with general infection and cure times,* Physical Review E **87**, 062816 (2013).

[51] V. Pisarenko and M. Rodkin, *Heavy-tailed distributions in disaster analysis,* Vol. 30 (Springer Science & Business Media, 2010).

[52] D. C. Dickson and H. R. Waters, *The distribution of the time to ruin in the classical risk model,* Astin Bulletin **32**, 299 (2002).

[53] S. Kostinski and A. Amir, *An elementary derivation of first and last return times of 1D random walks,* American Journal of Physics **84**, 57 (2016).

[54] Z. He and P. Van Mieghem, *The fastest spreader in SIS epidemics on networks,* The European Physical Journal B **91**, 77 (2018).

[55] L. Lü, D. Chen, X. Ren, Q. Zhang, Y. Zhang, and T. Zhou, *Vital nodes identification in complex networks,* Physics Reports **650**, 1 (2016).

[56] T. Zhou, J. Liu, W. Bai, G. Chen, and B. Wang, *Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity,* Physical Review E **74**, 056109 (2006).

[57] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, *Theory of rumour spreading in complex social networks,* Physica A **374**, 457 (2007).

[58] J. M. Hernández and P. Van Mieghem, *Classification of graph metrics,* Delft University of Technology, report20111111 (2011).

[59] T. Martin, X. Zhang, and M. Newman, *Localization and centrality in networks,* Physical Review E **90**, 052808 (2014).

[60] P. Van Mieghem, *Graph eigenvectors, fundamental weights and centrality metrics for nodes in networks,* arXiv preprint arXiv:1401.4580 (2014).

[61] A. Zeng and C. Zhang, *Ranking spreaders by decomposing complex networks,* Physics Letters A **377**, 1031 (2013).

[62] J. Liu, Z. Ren, and Q. Guo, *Ranking the spreading influence in complex networks,* Physica A **392**, 4154 (2013).

[63] F. Morone and H. A. Makse, *Influence maximization in complex networks through optimal percolation,* Nature (2015).

[64] P. Van Mieghem, K. Devriendt, and H. Cetinay, *Pseudoinverse of the Laplacian and best spreader node in a network,* Physical Review E **96**, 032311 (2017).

[65] M. Šikić, A. Lančić, N. Antulov-Fantulin, and H. Štefančić, *Epidemic centrality – Is there an underestimated epidemic impact of network peripheral nodes?* The European Physical Journal B **86**, 440 (2013).

[66] J. Liu, J. Lin, Q. Guo, and T. Zhou, *Locating influential nodes via dynamics-sensitive centrality,* Scientific Reports **6** (2016).

[67] P. Holme, *Three faces of node importance in network epidemiology: Exact results for small graphs,* Physical Review E **96**, 062305 (2017).

[68] B. Qu, C. Li, P. Van Mieghem, and H. Wang, *Ranking of nodal infection probability in susceptible-infected-susceptible epidemic,* Scientific Reports **7**, 9233 (2017).

[69] R. Van Der Hofstad, G. Hooghiemstra, and P. Van Mieghem, *The flooding time in random graphs,* Extremes **5**, 111 (2002).

[70] F. Ball and P. Donnelly, *Branching process approximation of epidemic models,* Theory of Probability & Its Applications **37**, 119 (1993).

[71] D. E. Knuth, *The Stanford GraphBase: a platform for combinatorial computing,* Vol. 37 (Addison-Wesley Reading, 1993).

[72] S. Milgram, *The small world problem,* Psychology Today **1**, 61 (1967).

[73] M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices,* Physical Review E **74**, 036104 (2006).

[74] O. Kariv and S. L. Hakimi, *An algorithmic approach to network location problems. i: The p-centers,* SIAM Journal on Applied Mathematics **37**, 513 (1979).

[75] T. M. Chan, *More algorithms for all-pairs shortest paths in weighted graphs,* SIAM Journal on Computing **39**, 2075 (2010).

[76] Z. He and P. Van Mieghem, *Optimal induced spreading of SIS epidemics in networks,* IEEE Transactions on Control of Network Systems **6**, 1344 (2018).

[77] S. Roy, M. Xue, and S. K. Das, *Security and discoverability of spread dynamics in cyber-physical networks,* IEEE Transactions on Parallel and Distributed Systems **23**, 1694 (2012).

[78] C. Nowzari, V. M. Preciado, and G. J. Pappas, *Analysis and control of epidemics: A survey of spreading processes on complex networks,* IEEE Control Systems **36**, 26 (2016).

[79] Y. Sun, L. Ma, A. Zeng, and W.-X. Wang, *Spreading to localized targets in complex networks,* Scientific Reports **6**, 38865 (2016).

[80] P. J. Roberts and C. J. Der, *Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer,* Oncogene **26**, 3291 (2007).

[81] W. Lu, S. Xu, and X. Yi, *Optimizing active cyber defense,* in *International Conference on Decision and Game Theory for Security* (Springer, 2013) pp. 206–225.

[82] A. Y. Lokhov and D. Saad, *Optimal deployment of resources for maximizing impact in spreading processes,* Proceedings of the National Academy of Sciences , 201614694 (2017).

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

[84] P. Van Mieghem and J. Omic, *In-homogeneous virus spread in networks,* arXiv preprint arXiv:1306.2588 (2013).

[85] B. Qu and H. Wang, *SIS epidemic spreading with heterogeneous infection rates,* IEEE Transactions on Network Science and Engineering **4**, 177 (2017).

[86] R. Storn and K. Price, *Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces,* Journal of Global Optimization **11**, 341 (1997).

[87] K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: a practical approach to global optimization* (Springer Science & Business Media, 2006).

[88] M. M. Ali and A. Törn, *Population set-based global optimization algorithms: some modifications and numerical studies,* Computers & Operations Research **31**, 1703 (2004).

[89] S. Lenhart and J. T. Workman, *Optimal control applied to biological models* (Crc Press, 2007).

[90] D. Liberzon, *Calculus of variations and optimal control theory: a concise introduction* (Princeton University Press, 2011).

[91] M. McAsey, L. Mou, and W. Han, *Convergence of the forward-backward sweep method in optimal control,* Computational Optimization and Applications **53**, 207 (2012).

[92] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control* (John Wiley & Sons, 2012).

[93] J. A. Primbs, V. Nevistić, and J. C. Doyle, *Nonlinear optimal control: A control lyapunov function and receding horizon perspective,* Asian Journal of Control **1**, 14 (1999).

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

[95] D. Lusseau and M. E. Newman, *Identifying the role that animals play in their social networks,* Proceedings of the Royal Society of London B: Biological Sciences **271**, S477 (2004).

[96] P. G. Doyle and J. L. Snell, *Random walks and electric networks,* Vol. 22 (Mathematical Association of America, 1984).

[97] I. Klickstein, A. Shirin, and F. Sorrentino, *Energy scaling of targeted optimal control of complex networks,* Nature Communications **8**, 15145 (2017).

[98] A. N. Bishop and I. Shames, *Link operations for slowing the spread of disease in complex networks,* EPL (Europhysics Letters) **95**, 18005 (2011).

[99] M. Zargham and V. Preciado, *Worst-case scenarios for greedy, centrality-based network protection strategies,* in *48th Annual Conference on Information Sciences and Systems (CISS)* (IEEE, 2014) pp. 1–6.

[100] R. Morton and K. H. Wickwire, *On the optimal control of a deterministic epidemic,* Advances in Applied Probability **6**, 622 (1974).

[101] G. A. Forster and C. A. Gilligan, *Optimizing the control of disease infestations at the landscape scale,* Proceedings of the National Academy of Sciences **104**, 4984 (2007).

[102] S. Eshghi, M. Khouzani, S. Sarkar, and S. Venkatesh, *Optimal patching in clustered epidemics of malware,* IEEE Transactions on Network (2015).

[103] L. Yang, X. Yang, and Y. Wu, *The impact of patch forwarding on the prevalence of computer virus: A theoretical assessment approach,* Applied Mathematical Modelling **43**, 110 (2017).

[104] Z. He and P. Van Mieghem, *Prevalence expansion in NIMFA,* Physica A: Statistical Mechanics and its Applications **540**, 123220 (2020).

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

[106] K. Devriendt and P. Van Mieghem, *Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality,* Physical Review E **96**, 052314 (2017).

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

[108] P. Van Mieghem, *The viral conductance of a network,* Computer Communications **35**, 1494 (2012).

[109] E. C. Titchmarsh, *The theory of functions* (Oxford University Press, 1968).

[110] G. Mercer and A. Roberts, *A centre manifold description of contaminant dispersion in channels with varying flow properties,* SIAM Journal on Applied Mathematics **50**, 1547 (1990).

[111] E. J. Hinch, *Perturbation methods* (Cambridge university press, 1991).

[112] D. B. Wilson, *Generating random spanning trees more quickly than the cover time,* in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (ACM, 1996) pp. 296–303.

[113] R. V. Solé and S. Valverde, *Information theory of complex networks: on evolution and architectural constraints,* in *Complex networks* (Springer, 2004) pp. 189–207.

[114] M. E. Newman, *Modularity and community structure in networks,* Proceedings of the national academy of sciences **103**, 8577 (2006).

[115] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications,* Bulletin of the American Mathematical Society **43**, 439 (2006).

[116] F. D. Malliaros and V. Megalooikonomou, *Expansion properties of large social graphs,* in *International Conference on Database Systems for Advanced Applications* (Springer, 2011) pp. 311–322.

[117] Z. He, P. Sun, and P. Van Mieghem, *Topological approach to measure network recoverability,* in *2019 11th International Workshop on Resilient Networks Design and Modeling (RNDM)* (IEEE, 2019).

[118] J. L. Marzo, S. G. Cosgaya, N. Skorin-Kapov, C. Scoglio, and H. Shakeri, *A study of the robustness of optical networks under massive failures,* Optical Switching and Networking **31**, 1 (2019).

[119] E. K. Çetinkaya and J. P. Sterbenz, *A taxonomy of network challenges,* in *2013 9th International Conference on the Design of Reliable Communication Networks (DRCN)* (IEEE, 2013) pp. 322–330.

[120] P. Cholda, A. Mykkeltveit, B. E. Helvik, O. J. Wittner, and A. Jajszczyk, *A survey of resilience differentiation frameworks in communication networks,* IEEE Communications Surveys & Tutorials **9**, 32 (2007).

[121] A. Pašić, R. Girão-Silva, B. Vass, T. Gomes, and P. Babarczi, *Fradir: A novel framework for disaster resilience,* in *2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM)* (IEEE, 2018) pp. 1–7.

[122] J. L. Marzo, E. Calle, S. G. Cosgaya, D. Rueda, and A. Mañosa, *On selecting the relevant metrics of network robustness,* in *2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM)* (IEEE, 2018) pp. 1–7.

[123] D. F. Rueda, E. Calle, and J. L. Marzo, *Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements,* Journal of Network and Systems Management **25**, 269 (2017).

[124] X. Pan and H. Wang, *Resilience of and recovery strategies for weighted networks,* PloS one **13**, e0203894 (2018).

[125] T. Afrin and N. Yodo, *A concise survey of advancements in recovery strategies for resilient complex networks,* Journal of Complex Networks (2018).

[126] W. Sun and A. Zeng, *Target recovery in complex networks,* The European Physical Journal B **90**, 10 (2017).

[127] J. Wang, C. Qiao, and H. Yu, *On progressive network recovery after a major disruption,* in *2011 IEEE INFOCOM Proceedings* (IEEE, 2011) pp. 1925–1933.

[128] K. Al Sabeh, M. Tornatore, and F. Dikbiyik, *Progressive network recovery in optical core networks,* in *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)* (IEEE, 2015) pp. 106–111.

[129] D. Z. Tootaghaj, H. Khamfroush, N. Bartolini, S. Ciavarella, S. Hayes, and T. La Porta, *Network recovery from massive failures under uncertain knowledge of damages,* in *2017 IFIP Networking Conference (IFIP Networking) and Workshops* (IEEE, 2017) pp. 1–9.

[130] V. Latora and M. Marchiori, *Efficient behavior of small-world networks,* Physical Review Letters **87**, 198701 (2001).

[131] S. Trajanovski, J. Martín-Hernández, W. Winterbach, and P. Van Mieghem, *Robustness envelopes of networks,* Journal of Complex Networks **1**, 44 (2013).

[132] P. Van Mieghem, *Data Communications Networking* (Purdue University Press, 2006).

[133] C. Li, H. Wang, W. De Haan, C. Stam, and P. Van Mieghem, *The correlation of metrics in complex networks with applications in functional brain networks,* Journal of Statistical Mechanics: Theory and Experiment **2011**, P11018 (2011).

[134] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, *The internet topology zoo,* IEEE Journal on Selected Areas in Communications **29**, 1765 (2011).

[135] Y. T. Woldeyohannes and Y. Jiang, *Measures for network structural dependency analysis,* IEEE Communications Letters **22**, 2052 (2018).

[136] W. Ellens, F. Spieksma, P. Van Mieghem, A. Jamakovic, and R. Kooij, *Effective graph resistance,* Linear Algebra and its Applications **435**, 2491 (2011).

[137] X. Wang, J. L. Dubbeldam, and P. Van Mieghem, *Kemeny's constant and the effective graph resistance,* Linear Algebra and its Applications **535**, 231 (2017).

[138] X. Wang, E. Pournaras, R. E. Kooij, and P. Van Mieghem, *Improving robustness of complex networks via the effective graph resistance,* The European Physical Journal B **87**, 221 (2014).

[139] A. Tizghadam and A. Leon-Garcia, *Autonomic traffic engineering for network robustness,* IEEE Journal on Selected Areas in Communications **28**, 39 (2010).

[140] H. R. TO and M. M. Barker, *White paper: European Transport Policy for 2010: time to decide,* (2001).

[141] M. Zhang and A. J. Pel, *Synchromodal hinterland freight transport: Model study for the port of Rotterdam,* Journal of Transport Geography **52**, 1 (2016).

[142] European Conference of Ministers of Transport, *Terminology on combined transport,* (1993).

[143] Logistiek, Topteam, *Partituur naar de top. Adviesrapport Topteam Logistiek,* (2011).

[144] B. van Riessen, R. R. Negenborn, and R. Dekker, *Synchromodal container transportation: an overview of current topics and research opportunities,* in *International conference on computational logistics* (Springer, 2015) pp. 386–397.

[145] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Complex networks: Structure and dynamics,* Physics Reports **424**, 175 (2006).

[146] Å. J. Holmgren, *A framework for vulnerability assessment of electric power systems,* in *Critical Infrastructure* (Springer, 2007) pp. 31–55.

[147] B. Berche, C. Von Ferber, T. Holovatch, and Y. Holovatch, *Resilience of public transport networks against attacks,* The European Physical Journal B **71**, 125 (2009).

[148] L.-G. Mattsson and E. Jenelius, *Vulnerability and resilience of transport systems – A discussion of recent research,* Transportation Research Part A: Policy and Practice **81**, 16 (2015).

[149] E. Rodríguez-Núñez and J. C. García-Palomares, *Measuring the vulnerability of public transport networks,* Journal of Transport Geography **35**, 50 (2014).

[150] H. Al-Deek and E. B. Emam, *New methodology for estimating reliability in transportation networks with degraded link capacities,* Journal of Intelligent Transportation Systems **10**, 117 (2006).

[151] J. L. Sullivan, D. C. Novak, L. Aultman-Hall, and D. M. Scott, *Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach,* Transportation Research Part A: Policy and Practice **44**, 323 (2010).

[152] A. Chen, H. Yang, H. K. Lo, and W. H. Tang, *Capacity reliability of a road network: an assessment methodology and numerical results,* Transportation Research Part B: Methodological **36**, 225 (2002).

[153] O. Cats, G. J. Koppenol, and M. Warnier, *Robustness assessment of link capacity reduction for complex networks: Application for public transport systems,* Reliability Engineering & System Safety **167**, 544 (2017).

[154] M. Zhang, B. Wiegmans, and L. Tavasszy, *Optimization of multimodal networks including environmental costs: A model and findings for transport policy,* Computers in Industry **64**, 136 (2013).

[155] P. Van Mieghem, *Interconnectivity structure of a general interdependent network,* Physical Review E **93**, 042305 (2016).

[156] P. Demeester, M. Gryseels, A. Autenrieth, C. Brianza, L. Castagna, G. Signorelli, R. Clemenfe, M. Ravera, A. Jajszczyk, D. Janukowicz, *et al.*, *Resilience in multilayer networks,* IEEE Communications Magazine **37**, 70 (1999).

[157] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Catastrophic cascade of failures in interdependent networks,* Nature **464**, 1025 (2010).

[158] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. Brazier, *A topological investigation of phase transitions of cascading failures in power grids,* Physica A: Statistical Mechanics and its Applications **415**, 273 (2014).

[159] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, *Navigability of interconnected networks under random failures,* Proceedings of the National Academy of Sciences **111**, 8351 (2014).

[160] A. A. Ganin, M. Kitsak, D. Marchese, J. M. Keisler, T. Seager, and I. Linkov, *Resilience and efficiency in transportation networks,* Science Advances **3**, e1701079 (2017).

[161] S. C. Dafermos and F. T. Sparrow, *The traffic assignment problem for a general network,* Journal of Research of the National Bureau of Standards B **73**, 91 (1969).

[162] T. G. Crainic and J. M. Rousseau, *Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem,* Transportation Research Part B: Methodological **20**, 225 (1986).

[163] S. Çolak, A. Lima, and M. C. González, *Understanding congested travel in urban areas,* Nature Communications **7**, 10793 (2016).

[164] F. Southworth and B. E. Peterson, *Intermodal and international freight network modeling,* Transportation Research Part C: Emerging Technologies **8**, 147 (2000).

[165] M. Patriksson, *The traffic assignment problem: models and methods* (Courier Dover Publications, 2015).

[166] R. Bellman, *On a routing problem,* Quarterly of Applied Mathematics **16**, 87 (1958).

[167] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks,* Mathematics of Operations Research **29**, 961 (2004).

[168] J. G. Wardrop, *Road paper. Some theoretical aspects of road traffic research.* Proceedings of the institution of civil engineers **1**, 325 (1952).

[169] M. Fukushima, *A modified Frank-Wolfe algorithm for solving the traffic assignment problem,* Transportation Research Part B: Methodological **18**, 169 (1984).

[170] J. N. Prashker and S. Bekhor, *Some observations on stochastic user equilibrium and system optimum of traffic assignment,* Transportation Research Part B: Methodological **34**, 277 (2000).

[171] P. A. Chen and D. Kempe, *Altruism, selfishness, and spite in traffic routing,* in *Proceedings of the 9th ACM conference on Electronic commerce* (ACM, 2008) pp. 140–149.

[172] R. L. Tobin and T. L. Friesz, *Sensitivity analysis for equilibrium network flow,* Transportation Science **22**, 242 (1988).

[173] M. Patriksson, *Sensitivity analysis of traffic equilibria,* Transportation Science **38**, 258 (2004).

[174] W. van Dam, *Robustness analysis of the Dutch synchromodal freight transport network: Simulating disruptions on a macroscopic graph model,* (2017).

[175] O. Cats, *Topological evolution of a metropolitan rail transport network: The case of Stockholm,* Journal of Transport Geography **62**, 172 (2017).

[176] D. Braess, *Über ein paradoxon aus der verkehrsplanung,* Unternehmensforschung **12**, 258 (1968).

[177] L. Zhang, G. Zeng, D. Li, H.-J. Huang, H. E. Stanley, and S. Havlin, *Scale-free resilience of real traffic jams,* Proceedings of the National Academy of Sciences **116**, 8673 (2019).

[178] I. Simonsen, K. A. Eriksen, S. Maslov, and K. Sneppen, *Diffusion on complex networks: A way to probe their large-scale topological structures,* Physica A: Statistical Mechanics and its Applications **336**, 163 (2004).

[179] R. Olfati-Saber, J. A. Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems,* Proceedings of the IEEE **95**, 215 (2007).

[180] Z. Li, Z. Duan, G. Chen, and L. Huang, *Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint,* IEEE Transactions on Circuits and Systems I: Regular Papers **57**, 213 (2010).

[181] A. Ghosh and S. Boyd, *Growing well-connected graphs,* in *2006 45th IEEE Conference on Decision and Control* (IEEE, 2006) pp. 6605–6611.

[182] Y. Kim and M. Mesbahi, *On maximizing the second smallest eigenvalue of a state-dependent graph Laplacian,* IEEE Transactions on Automatic Control **51**, 116 (2006).

[183] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph,* SIAM Review **46**, 667 (2004).

[184] A. Sydney, C. Scoglio, and D. Gruenbacher, *Optimizing algebraic connectivity by edge rewiring,* Applied Mathematics and Computation **219**, 5465 (2013).

[185] K. Ogiwara, T. Fukami, and N. Takahashi, *Maximizing algebraic connectivity in the space of graphs with a fixed number of vertices and edges,* IEEE Transactions on Control of Network Systems **4**, 359 (2017).

[186] M. J. Alenazi, E. K. Cetinkaya, and J. P. Sterbenz, *Network design and optimisation based on cost and algebraic connectivity,* in *2013 5th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (IEEE, 2013) pp. 193–200.

[187] G. Li, Z. F. Hao, H. Huang, and H. Wei, *Maximizing algebraic connectivity via minimum degree and maximum distance,* IEEE Access **6**, 41249 (2018).

[188] C. Li, Z. Qu, A. Das, and F. Lewis, *Cooperative control with improvable network connectivity,* in *American Control Conference (ACC), 2010* (IEEE, 2010) pp. 87–92.

[189] M. M. Asadi, M. Khosravi, A. G. Aghdam, and S. Blouin, *Expected convergence rate to consensus in asymmetric networks: Analysis and distributed estimation,* IEEE Transactions on Systems, Man, and Cybernetics: Systems (2017).

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

[191] G. W. Stewart, *Matrix perturbation theory,* (1990).

[192] C. Li and Z. Qu, *Distributed estimation of algebraic connectivity of directed networks,* Systems & Control Letters **62**, 517 (2013).

[193] A. Quarteroni, R. Sacco, and F. Saleri, *Matematica numerica* (Springer Science & Business Media, 2008).

[194] A. Milanese, J. Sun, and T. Nishikawa, *Approximating spectral impact of structural perturbations in large networks,* Physical Review E **81**, 046112 (2010).

[195] L. Lovász, *Submodular functions and convexity,* in *Mathematical Programming: The State of the Art* (Springer, 1983) pp. 235–257.

[196] M. Feldman, C. Harshaw, and A. Karbasi, *Greed is good: Near-optimal submodular maximization via greedy optimization,* arXiv preprint arXiv:1704.01652 (2017).

[197] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, *An analysis of approximations for maximizing submodular set functions—I,* Mathematical programming **14**, 265 (1978).

[198] Y. Liu, Z. Zhang, E. K. Chong, and A. Pezeshki, *Performance bounds with curvature for batched greedy optimization,* Journal of Optimization Theory and Applications **177**, 535 (2018).

[199] R. Merris, *Laplacian matrices of graphs: A survey,* Linear Algebra and its Applications **197**, 143 (1994).

[200] G. Gupta, S. Pequito, and P. Bogdan, *Approximate submodular functions and performance guarantees,* arXiv preprint arXiv:1806.06323 (2018).

[201] T. Horel and Y. Singer, *Maximization of approximately submodular functions,* in *Advances in Neural Information Processing Systems* (2016) pp. 3045–3053.

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

[203] T. Kato, *Perturbation theory for linear operators,* Vol. 132 (Springer Science & Business Media, 2013).

[204] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems,* in *Recent Advances in Iterative Methods* (Springer, 1994) pp. 149–164.

[205] F. L. Bauer and C. T. Fike, *Norms and exclusion theorems,* Numerische Mathematik **2**, 137 (1960).

[206] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression diagnostics: Identifying influential data and sources of collinearity,* Vol. 571 (John Wiley & Sons, 2005).

[207] J. S. Coleman, *Introduction to mathematical sociology.* (New York: Free Press, 1964).

[208] O. Jahn, R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses, *System-optimal routing of traffic flows with user constraints in networks with congestion,* Operations Research **53**, 600 (2005).

[209] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, *The structure of the nervous system of the nematode Caenorhabditis elegans,* Philos Trans R Soc Lond B Biol Sci **314**, 1 (1986).

[210] H. Zhidong and H. Wenjun, *Extremal optimization approach to joint routing and scheduling for industrial wireless networks,* in *2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems* (IEEE, 2013) pp. 427–428.

[211] F. D. Malliaros and M. Vazirgiannis, *Clustering and community detection in directed networks: A survey,* Physics Reports **533**, 95 (2013).

[212] N. Blenn and P. Van Mieghem, *Are human interactivity times lognormal?* arXiv preprint arXiv:1607.02952 (2016).

[213] R. Lawrence, *The lognormal distribution of the duration of strikes,* Journal of the Royal Statistical Society. Series A (General) , 464 (1984).

[214] P. E. Sartwell *et al.*, *The distribution of incubation periods of infectious diseases.* American Journal of Hygiene **51**, 310 (1950).

[215] P. Philippe, *Sartwell's incubation period model revisited in the light of dynamic modeling,* Journal of Clinical Epidemiology **47**, 419 (1994).

[216] B. Ottino-Loffler, J. G. Scott, and S. H. Strogatz, *Evolutionary dynamics of incubation periods,* eLife **6**, e30212 (2017).

[217] T. H. Summers, F. L. Cortesi, and J. Lygeros, *On submodularity and controllability in complex dynamical networks,* IEEE Transactions on Control of Network Systems **3**, 91 (2015).

[218] B. Recht, *A tour of reinforcement learning: The view from continuous control,* Annual Review of Control, Robotics, and Autonomous Systems **2**, 253 (2019).

[219] F. L. Lewis and D. Vrabie, *Reinforcement learning and adaptive dynamic programming for feedback control,* IEEE Circuits and Systems Magazine **9**, 32 (2009).

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

[221] J. S. Andrade Jr, E. Oliveira, A. Moreira, and H. J. Herrmann, *Fracturing the optimal paths,* Physical Review Letters **103**, 225503 (2009).

[222] D. Guo, S. Trajanovski, R. van de Bovenkamp, H. Wang, and P. Van Mieghem, *Epidemic threshold and topological structure of susceptible-infectious-susceptible epidemics in adaptive networks,* Physical Review E **88**, 042802 (2013).

[223] E. A. Bender, *Asymptotic methods in enumeration,* SIAM Review **16**, 485 (1974).

ACKNOWLEDGEMENTS

First of all, I owe my great gratitude to my promoter, Prof. Piet Van Mieghem. It has been my honor to have been in NAS group and working with him. Piet has impressed me with his passion and his academic rigor in research approaches and findings. Piet has always emphasized the importance of telling an academic story in English and reminded me of the KISS principle (Keep It Short & Simple). I keep it in mind and will regard it as a dogma for further career.

During my Ph.D. studies, I have been fortunate to work together and collaborate with some distinguished researchers and master students, Peng Sun, Wirdmer van Dam, Kumar Navneet, Dr. Kai Yuan, and Prof. Robert E. Kooij. I am very grateful for the experience for exploring original issues with them. In addition, I would like to express my gratitude to the committee members of my thesis defense for their reviewing and efforts on my thesis.

Next, I would like to thank my office mates, Dr. Xiangrong Wang, Dr. Marcus Märtens and Dr. Hale Çetinay. Hale is the office mate with me for the longest time. Her talks and gossips have made my office life enjoyable. Xiangrong has helped me a lot as a kind Chinese senior sister. I also enjoyed talking with Misa Taguchi about different cultures in various languages. I also thank my current office mates. Peng Sun, who likes sharing his Hifi experience with me, is essentially a good person. Ivan Jokić often discusses his new findings with me. His working efficiency and gentleness impressed me. Gabriel Budel speaks Chinese well and kindly helps me by his excellent language talents. Gabriel also contributes to the Dutch translation parts in this thesis. Fenghua Wang always disturbs my study by her vivacity and requests for help. I like the atmosphere of my office and wish everyone success.

Many thanks also go to other colleagues in NAS group, Dr. Edgar van Boven, Dr. Remco Litjens, Dr. Eric Smeitink, Rogier Noldus, Dr. Jaron Sanders, Dr. Niels van Adrichem, Dr. Norbert Blenn, Dr. Farabi Iqbal, Dr. Jil Meier, Karel Devriendt, Bastian Prasse, Dr. Chuan Lin, Maria Raftopoulou, Qingfeng Tong, and Massimo Achterberg, Albert Senen-Cerda. I owe special thanks to Dr. Qiang Liu, Long Ma and Dr. Luxing Yang. I have the most memorable experience with them in both academic study and spare-time life.

As an aloof fish, I prefer to stay at home for relaxation and entertainment. I want to mention and recommend group Nogizaka46. The TV programmes of its members Sayuri Matsumura and Momoko Ozono have a therapeutic effect when I felt lost. I wish that Mai Shiraishi, who will also graduate in 2020, has a better career.

My final and warmest thanks to my parents for their emotional and financial supports. I will try to fulfill their wills to have a marriage as soon as possible.

Zhidong He
Delft, March 2020

A

APPENDIX FOR CHAPTER 2

A.1. DETERMINATION OF THE METASTABLE STATE AND THE STABILITY TIME

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

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

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

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

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

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

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

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

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

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

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

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

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

In the above definition, we introduce the prevalence

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

subject to the condition that the process does not die out, where $\Pr[I(t) = 0]$ is the extinction probability. The prevalence $\overline{y}(t)$ excluding early extinction, as illustrated in Figure A.1, tends to stay almost constant instead of decaying as the prevalence $y(t)$ after reaching the extremum. We consider that the metastable state starts when the prevalence $\overline{y}(t)$ stays almost constant. Actually, the prevalence $\overline{y}(t)$ excluding early extinction is a monotonically increasing function, which only stays constant when $t \to \infty$, as follows from general Markov theory [38]. The prescribed stringent parameter ϵ can be determined as a small value. Definition 1(e) is also consistent with the definition of

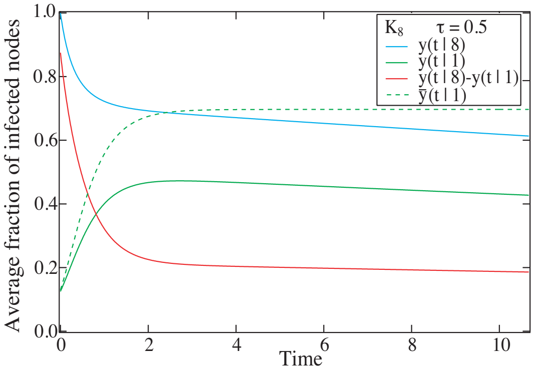


Figure A.1: The exact prevalences $y(t|1)$ started from one infected node and the exact prevalence $y(t|8)$ started from all infected nodes in a complete graph K_8 with the effective infection rate $\tau = 0.5$. The red line represents the difference between the two prevalences $y(t|1)$ and $y(t|8)$. The green dash line represents the prevalence $\overline{y}(t|1)$ excluding early extinction probability.

the quasi-stationary state, which leads to the almost steady average number of infected nodes without extinction realizations.

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

A.2. SIMULATION FOR A SIS PROCESS ON NETWORKS

There are two kinds of events in the SIS process which are infection events and curing events. All the events are marked on a same timeline and are handled by the order of their time after the beginning of the simulation. We denote by $\Omega_{n,1}(t)$ the infection event that node n becomes infected at time t, and $\Omega_{n,0}(t)$ the curing event that node n becomes cured at time t. The process of SSIS (Simulation for SIS epidemics) is described by Algorithm 1.1.

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

A.3. THE GENERATING FUNCTION OF THE HITTING TIME

We consider the hitting time when the process first reaches the state with n infected nodes starting from one initial spreader. Let Q_n be the infinitesimal generator of the modified Markovian SIS as

$$Q_n = \begin{bmatrix} -q_{1;1} & q_{1;2} & 0 & \cdots & 0 \\ q_{2;1} & -q_{2;2} & q_{2;3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & q_{n-1;n-2} & -q_{n-1;n-1} & q_{n-1;n} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

We define $\widetilde{Q}_i$ as the sub-matrix of the first i rows and first i column of Q_n, and define $\widetilde{I}_i$ as the $i \times i$ unit matrix. We denote by $T_{i,i+j}$ the hitting time of state $i+j$ starting from state i. The Laplace transform of $T_{i,i+j}$ is $\varphi_{i,i+j}(z) = E[e^{-zT_{i,i+j}}]$. The sojourn time $T_{i,i}$ in the state i follows an exponential with rate $q_{i;i}$, and its corresponding Laplace transform is $\varphi_{i,i}(z) = \frac{q_{i;i}}{q_{i;i}+z}$.

In the continuous Markov chain, the process stays state i for a sojourn time, then jumps to $i+1$ with probability $\frac{q_{i;i+1}}{q_{i;i}}$, to $i-1$ with probability $\frac{q_{i;i-1}}{q_{i;i}}$, $(2 \le i < n)$. Thus, the probability of the hitting time for each jump follows

$$\Pr[T_{i,i+1} < t] = \frac{q_{i;i+1}}{q_{i;i}} \Pr[T_{i,i} < t] + \frac{q_{i;i-1}}{q_{i;i}} \Pr[T_{i,i} + T_{i-1,i+1} < t] \tag{A.3}$$

The Laplace transform of the above equation (A.3) reads

$$\varphi_{i,i+1}(z) = \frac{q_{i;i}}{q_{i;i}+z}\frac{q_{i;i+1}}{q_{i;i}} + \frac{q_{i;i}}{q_{i;i}+z}\frac{q_{i;i-1}}{q_{i;i}}\varphi_{i-1,i+1}(z) \tag{A.4}$$

Since $T_{1,j}$ and $T_{j,i}$ are independent random variables, we have $\varphi_{1,i}(z) = \varphi_{1,j}(z)\varphi_{j,i}(z)$ for $1 < j < i$. Then, dividing (A.4) by $\varphi_{1,i+1}(z)$ on both sides gives

$$\frac{q_{i;i}+z}{\varphi_{1,i}(z)} = \frac{q_{i;i+1}}{\varphi_{1,i+1}(z)} + \frac{q_{i;i-1}}{\varphi_{1,i-1}(z)} \tag{A.5}$$

We further define

$$g_{1,1}(z) = 1; \quad g_{1,i}(z) = \frac{\prod_{j=1}^{i-1} q_{j,j+1}}{\varphi_{1,i}(z)}, 2 \le j \le i-1$$

Substituting $g_{1,i}(z)$ to (A.5) yields

$$g_{1,i+1}(z) = (q_{i;i}+z)g_{1,i}(z) - q_{i;i-1}q_{i-1;i}g_{1,i-1}(z) \tag{A.6}$$

From (A.6), the induction reads

$$\begin{aligned} g_{1,2} &= \frac{q_{1,2}}{\varphi_{1,2}(z)} = z + q_{1,1} = \det|z\widetilde{I}_1 - \widetilde{Q}_1| \\ g_{1,3} &= (q_{2;2}+z)g_{1,2}(z) - q_{2;1}q_{1;2}g_{1,1}(z) = \det|z\widetilde{I}_2 - \widetilde{Q}_2| \\ &\cdots \\ g_{1,n} &= \det|z\widetilde{I}_{n-1} - \widetilde{Q}_{n-1}| \end{aligned}$$

Meanwhile, let η_i be the eigenvalues of $-\widetilde{Q}_{n-1}$. For $z = 0$, we have $\varphi_{1,i}(0) = 1$ and

$$\prod_{i=1}^{n-1} q_{i,i+1} = g_{1,n}(1)\varphi_{1,n}(0) = \det|-\widetilde{Q}_{n-1}| = \prod_{i=1}^{n-1} \eta_i$$

Since that $\det|z\widetilde{I}_n - (-Q_n)| = (-1)^{n+n} z \det|z\widetilde{I}_{n-1} - (-\widetilde{Q}_{n-1})| = z\prod_{i=1}^{n-1}(z-\eta_i)$, the eigenvalues η_i are also the non-zero eigenvalues of $-Q_n$.

Therefore, we can obtain the conclusion that the generating function of the hitting time follows

$$\varphi_{1,n}(z) = \frac{\prod_{i=1}^{n-1} q_{i,i+1}}{g_{1,n}(z)} = \frac{\prod_{i=1}^{n-1} \eta_i}{\det(z\widetilde{I}_{n-1} - \widetilde{Q}_{n-1})} = \prod_{i=1}^{n-1} \frac{\eta_i}{\eta_i + z}, \tag{A.7}$$

where η_i are the non-zero eigenvalues of $-Q_n$.

Algorithm 1.1 Simulation for SIS epidemics

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

B

Appendix for Chapter 4

B.1. Proofs of Lemma and Theorem

B.1.1. Proof for Lemma 4.1

Proof: According to the NIMFA equation (4.2), we have the steady-state equation

$$\Big(\mathrm{diag}(1-v_{i\infty})A\mathrm{diag}(\beta_j)-I\Big)\mathbf{v}_\infty=0 \tag{B.1}$$

By differentiation with respect to β_j, we obtain

$$\begin{cases} \displaystyle\sum_{k=1}^{N} a_{ik}\beta_k\frac{\partial v_k}{\partial \beta_i}-\frac{1}{(1-v_{i\infty})^2}\frac{\partial v_{i\infty}}{\partial \beta_i}=0 & \text{if } i=j \\ \displaystyle\sum_{k=1}^{N} a_{jk}\beta_k\frac{\partial v_k}{\partial \beta_i}-\frac{1}{(1-v_{j\infty})^2}\frac{\partial v_{j\infty}}{\partial \beta_i}+a_{ji}v_{i\infty}=0 & \text{if } i\neq j \end{cases}$$

Written in matrix form, we have

$$\Big(A\mathrm{diag}(\beta_i)-\mathrm{diag}\big((1-v_{i\infty})^{-2}\big)\Big)T_1+A\mathrm{diag}(\mathbf{v}_\infty)=0 \tag{B.2}$$

where the element of the matrix T_1 in the k-th row and the q-th column is $T_{1(kq)}=\frac{\partial v_{k\infty}}{\partial \beta_q}$. For the matrix $M_1:=A\mathrm{diag}(\beta_i)-\mathrm{diag}\big((1-v_{i\infty})^{-2}\big)$, it has been proved that all entries in M_1^{-1} are non-positive [84], which implies that all entries in $T_1=M_1^{-1}(-A\mathrm{diag}(\mathbf{v}_\infty))$ are non-negative. ☐

B.1.2. Proof for Theorem 4.1

Proof: By differentiation of (B.2) with respect to β_j, we obtain

$$\begin{cases} \displaystyle\sum_{k=1}^{N} a_{ik}\beta_k\frac{\partial^2 v_k}{\partial {\beta_i}^2}-\frac{1}{(1-v_{i\infty})^2}\frac{\partial^2 v_{i\infty}}{\partial {\beta_i}^2}-\frac{2}{(1-v_{i\infty})^3}\left(\frac{\partial v_{i\infty}}{\partial \beta_i}\right)^2=0 & \text{if } i=j \\ \displaystyle\sum_{k=1}^{N} a_{jk}\beta_k\frac{\partial^2 v_k}{\partial {\beta_i}^2}-\frac{1}{(1-v_{j\infty})^2}\frac{\partial^2 v_{j\infty}}{\partial {\beta_i}^2}-\frac{2}{(1-v_{j\infty})^3}\left(\frac{\partial v_{j\infty}}{\partial \beta_i}\right)^2+2a_{ji}\frac{\partial v_{i\infty}}{\partial \beta_i}=0 & \text{if } i\neq j \end{cases}$$

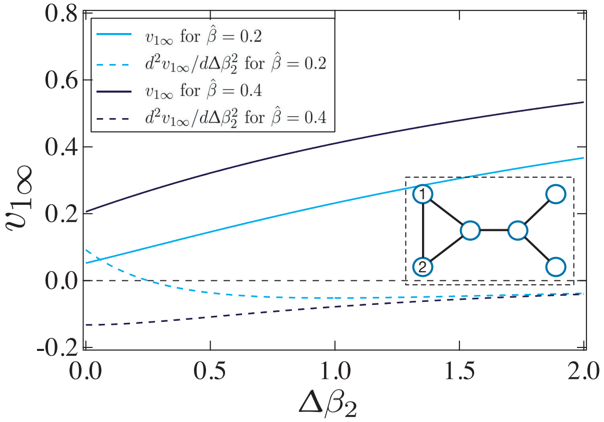


Figure B.1: The infection probability $v^{(3)}_{1\infty}$ and its second order partial derivative $\partial^2 v^{(3)}_{1\infty}/\partial\Delta\beta_2^2$ as the function of $\Delta\beta_2$ with different original infection rate $\hat{\beta}$ in the example network G_6.

Written in matrix form, we have

$$\left(A\mathrm{diag}(\beta_i) - \mathrm{diag}\frac{1}{(1-v_{i\infty})^2}\right) T_2 + 2A\mathrm{diag}\left(\frac{\partial v_{i\infty}}{\partial \beta_i}\right) - 2\mathrm{diag}\left((1-v_{i\infty})^{-3}\right) T_1^{(2)} = 0 \tag{B.3}$$

where the element of the matrix T_2 in the k-th row and the q-th column is $T_{2(kq)} = \frac{\partial^2 v_{k\infty}}{\partial \beta_q{}^2}$, and $T^{(2)}_{1(kq)} = \left(\frac{\partial v_{k\infty}}{\partial \beta_q}\right)^2$ for the matrix $T_1^{(2)}$. In the matrix $M_2 := 2A\mathrm{diag}\left(\frac{\partial v_{i\infty}}{\partial \beta_i}\right) - 2\mathrm{diag}\left((1-v_{i\infty})^{(-3)}\right) T_1^{(2)}$, the entries follows

$$M_{2(ij)} = 2\frac{\partial v_{i\infty}}{\partial \beta_j}\left(a_{ij} - (1-v_{i\infty})^{-3}\frac{\partial v_{i\infty}}{\partial \beta_j}\right). \tag{B.4}$$

Invoking that $\frac{\partial v_{i\infty}}{\partial \beta_i}$ is always non-negative, the entries $M_{2(ij)}$ in M_2 could be non-positive only when $a_{ij} = 0$. Unfortunately, the sign of $\frac{\partial^2 v_{i\infty}}{\partial \beta_j{}^2}$ cannot be determined and could be positive. Therefore, the infection probability $v_{i\infty}$ is not always concave to β_j. We present a counterexample in Fig. B.1. ☐

B.2. DIFFERENTIAL EVOLUTION ALGORITHM

We propose the Differential Evolution algorithm to solve the static induced SIS spreading problem. The implementation is as follow.

Population generation: The jth vector of the population at the kth generation is denoted as $\boldsymbol{\Delta\beta}_j(k) = \{\Delta\beta_{1,j}(k), \Delta\beta_{2,j}(k), \ldots, \Delta\beta_{N,j}(k)\}$. The initial population vectors are generated considering the constraints on the variables. We choose the initial generation as $\Delta\beta_{i,j}(1) = \kappa\Delta\overline{\beta}_i$, where κ is a uniformly distributed random number on interval $[0,1]$.

Mutation: Mutation is a change or perturbation with a random element. We choose three different vectors with indices $r_1, r_2, r_3 \in \{1, 2, \ldots, N_p\}$ and construct the mutated vector $\boldsymbol{\Delta\beta}_j(k+1)$ at the $(k+1)$th generation as $\boldsymbol{\Delta\beta}_j(k+1) = \boldsymbol{\Delta\beta}_{r_1}(k) + F(\boldsymbol{\Delta\beta}_{r_2}(k) - \boldsymbol{\beta}_{r_3}(k))$, where F is a uniformly distributed random number on interval $[F_{min}, F_{max}]$.

Crossover: Crossover is to enhance the potential diversity of the population, which obeys

$$\Delta\beta_{i,j}(k+1) = \begin{cases} \Delta\beta_{i,j}(k+1) & \text{if } \kappa \leq R \\ \Delta\beta_{i,j}(k) & \text{otherwise} \end{cases}$$

where R is the crossover rate, which is a prescribe parameter of the algorithm.

Selection: DE uses the greedy strategy to choose the better vector to be the population in the next generation. The selection operation is described as

$$\boldsymbol{\Delta\beta}_j(k+1) = \begin{cases} \boldsymbol{\Delta\beta}_j(k+1) & \text{if } J(\boldsymbol{\Delta\beta}_j(k+1)) \leq J(\boldsymbol{\Delta\beta}_j(k)) \\ \boldsymbol{\Delta\beta}_j(k) & \text{otherwise} \end{cases}$$

The process of the Differential Evolution algorithm for the static induced spreading problem is presented in Algorithm 2.1.

Algorithm 2.1 Differential Evolution algorithm

```
Inputs:
    $A$, $\beta$, $M$, $K$, $\epsilon$
Initialization:
    Set $k \leftarrow 1$
    Generate initial populations $\boldsymbol{\Delta\beta}_j(1), j \in \{1,2,\ldots,N_P\}$
for $k = 1$ to $K$ do
    for $j = 1$ to $N_P$ do
        Select randomly $r_1 \neq r_2 \neq r_3$ with $r_1, r_2, r_3 \in \{1,2,\ldots,N_P\} : j_{rand} = randint(1,N)$
        for $i = 1$ to $N$ do
            if $rand_j(0,1) < R$ or $j = j_{rand}$ then
                $\Delta\beta_{i,j}(k+1) = \Delta\beta_{i,r_1}(k) + F(\Delta\beta_{i,r_2}(k) - \Delta\beta_{i,r_3}(k))$
            else
                $\Delta\beta_{i,j}(k+1) = \Delta\beta_{i,j}(k)$
            end if
        end for
        if $J(\boldsymbol{\Delta\beta}_j(k+1)) \leq J(\boldsymbol{\Delta\beta}_j(k))$ then
            $\boldsymbol{\Delta\beta}_j(k+1) \leftarrow \boldsymbol{\Delta\beta}_j(k+1)$
        else
            $\boldsymbol{\Delta\beta}_j(k+1) \leftarrow \boldsymbol{\Delta\beta}_j(k)$
        end if
    end for
    $k \leftarrow k+1$
end for
```

C

APPENDIX FOR CHAPTER 5

C.1. RECURRENCE OF THE COEFFICIENTS $c_j(k)$ IN THE EXPANSION OF $v_{i\infty}(\tau)$

The steady-state infection probability $v_{i\infty}(\tau)$ of node i in SIS epidemics with the effective infection rate $\tau \downarrow \tau_c^{(1)}$ follows

$$v_{i\infty}(\tau) = \sum_{j=1}^{\infty} \sum_{k=1}^{N} c_j(k)(x_k)_i((\tau_c^{(1)})^{-1} - \tau^{-1})^j \tag{C.1}$$

where all coefficients $c_j(k)$ for the non-trivial solution $v_{i\infty}(\tau)$ are determined in a recursive way as follows [108].

Defining $T(m,l,k) = \sum_{q=1}^{N}(x_m)_q(x_l)_q(x_k)_q$, the coefficients $c_j(m)$ obey, for $m > 1$ and $j > 2$, the recursion

$$\begin{aligned} c_j(m) = {} & \frac{c_{j-1}(m)}{\lambda_1 - \lambda_m}\{1 - c_1(1)(\lambda_1 + \lambda_m)T(m,m,1)\} - \frac{c_1(1)}{\lambda_1 - \lambda_m} \sum_{k=1;k\neq m}^{N} (\lambda_1 + \lambda_k)c_{j-1}(k)T(m,k,1) \\ & - \frac{1}{\lambda_1 - \lambda_m} \sum_{n=2}^{j-2} \sum_{l=1}^{N} \sum_{k=1}^{N} c_{j-n}(l)c_n(k)\lambda_k T(m,l,k) \end{aligned} \tag{C.2}$$

while, for $j = 2$ and $m > 1$,

$$c_2(m) = -\frac{1}{\lambda_1 - \lambda_m} \frac{T(m,1,1)}{\lambda_1 T^2(1,1,1)} \tag{C.3}$$

and $c_1(m) = 0$. For $m = 1$, there holds that $c_1(1) = (\lambda_1 \sum_{j=1}^{N}(x_1)_j^3)^{-1}$ and for $j = 1$, the coefficients $c_j(1)$ satisfy the recursion

$$c_j(1) = -\frac{1}{\lambda_1 T(1,1,1)} \sum_{k=2}^{N} (\lambda_1 + \lambda_k)c_j(k)T(1,1,k) - \sum_{n=2}^{j-1} \sum_{l=1}^{N} \sum_{k=1}^{N} c_{j+1-n}(l)c_n(k)\lambda_k T(1,l,k) \tag{C.4}$$

C.2. Proof for Lemma 5.1

Proof: Without loss of generality, we assume that the convergence order of the elements $|(\boldsymbol{\alpha}_j)_i|$ for $i = 1,\dots,N$ are different, and the coefficient $|(\boldsymbol{\alpha}_j)_k|$ has the maximum convergence order, i.e., $\lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_k|^{\frac{1}{j}} = \max_{i\in N}\{\lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_i|^{\frac{1}{j}}\}$. Then there exists a critical order j_c such that $|(\boldsymbol{\alpha}_j)_k| = \max_{i\in N}|(\boldsymbol{\alpha}_j)_i|$ for $j > j_c$. The absolute coefficients $|b_j|$ follows

$$\begin{aligned}\lim_{j\to\infty}|b_j|^{\frac{1}{j}} &= \lim_{j\to\infty}\left|\frac{1}{N}u^T\boldsymbol{\alpha}_j\right|^{\frac{1}{j}} \le \lim_{j\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}|(\boldsymbol{\alpha}_j)_i|\right)^{\frac{1}{j}}\\ &= \lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_k|^{\frac{1}{j}}\left(\frac{1}{N}+\frac{1}{N}\sum_{i=1,i\neq k}^{N}\frac{|(\boldsymbol{\alpha}_j)_i|}{|(\boldsymbol{\alpha}_j)_k|}\right)^{\frac{1}{j}} = \lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_k|^{\frac{1}{j}}\end{aligned} \tag{C.5}$$

Meanwhile, the norm of the coefficients $||\boldsymbol{\alpha}_j||$ follows

$$\lim_{j\to\infty}||\boldsymbol{\alpha}_j||^{\frac{1}{j}} = \lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_k|^{\frac{1}{j}}\left(1+\sum_{i=1,i\neq k}^{N}\frac{|(\boldsymbol{\alpha}_j)_i|^2}{|(\boldsymbol{\alpha}_j)_k|^2}\right)^{\frac{1}{2j}} = \lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_k|^{\frac{1}{j}} \tag{C.6}$$

Hence, we obtain the relation that $\lim_{j\to\infty}|b_j|^{\frac{1}{j}} \le \lim_{j\to\infty}||\boldsymbol{\alpha}_j||^{\frac{1}{j}}$.

Invoking that the coefficient $\boldsymbol{\alpha}_j = X\boldsymbol{c}_j$ and supposing

$$\lim_{j\to\infty}|(\boldsymbol{c}_j)_m|^{\frac{1}{j}} = \max_{i\in N}\{\lim_{j\to\infty}|(\boldsymbol{c}_j)_i|^{\frac{1}{j}}\}$$

, we can obtain that

$$\lim_{j\to\infty}||\boldsymbol{\alpha}_j||^{\frac{1}{j}} = \lim_{j\to\infty}||\boldsymbol{c}_j||^{\frac{1}{j}} = \lim_{j\to\infty}|(\boldsymbol{c}_j)_m|^{\frac{1}{j}} = \lim_{j\to\infty}|(\boldsymbol{\alpha}_j)_k|^{\frac{1}{j}} \tag{C.7}$$

which implies that $||\boldsymbol{\alpha}_j||$ and $||\boldsymbol{c}_j||$ have the same convergence order. ☐

C.3. A lower bound of the radius of convergence R

We hereby heuristically propose a lower bound of the radius of convergence R. We rewrite (5.12) as

$$(\lambda_1 I - \Lambda)\boldsymbol{c}_j = X^T\boldsymbol{\alpha}_{j-1} - X^T\sum_{k=1}^{j-1} diag(\boldsymbol{\alpha}_k)A\boldsymbol{\alpha}_{j-1-k} \tag{C.8}$$

and define the vector $\tilde{\boldsymbol{c}}_j$ with $(\tilde{\boldsymbol{c}}_j)_1 = 0$ and $(\tilde{\boldsymbol{c}}_j)_m = (\boldsymbol{c}_j)_m$ for $m > 2$ and $j > 2$. We suppose that the norm of the coefficients $||\tilde{\boldsymbol{c}}_j||$ approximates $||\boldsymbol{c}_j||$ well for a large network size N. Further denoting the matrix $S := diag(0, \frac{1}{\lambda_1-\lambda_2},\dots,\frac{1}{\lambda_1-\lambda_N})$ and invoking $||\boldsymbol{\alpha}_j|| = ||\boldsymbol{c}_j||$, we can obtain the norm of coefficients $||\boldsymbol{\alpha}_j||$ follows

$$\begin{aligned}
||\boldsymbol{\alpha}_j|| \approx ||\tilde{\boldsymbol{c}}_j|| &= ||SX^T\boldsymbol{\alpha}_{j-1} - SX^T\sum_{k=1}^{j-1} diag(\boldsymbol{\alpha}_k)A\boldsymbol{\alpha}_{j-1-k}|| \\
&\leq ||S\boldsymbol{\alpha}_{j-1}|| + ||S\sum_{k=1}^{j-1} diag(\boldsymbol{\alpha}_k)A\boldsymbol{\alpha}_{j-1-k}|| \\
&\leq ||S||\cdot||\boldsymbol{\alpha}_{j-1}|| + ||S||\sum_{k=1}^{j-1} ||diag(\boldsymbol{\alpha}_k)||\cdot||A||\cdot||\boldsymbol{\alpha}_{j-1-k}|| \\
&\leq \frac{1}{\lambda_1-\lambda_2}||\boldsymbol{\alpha}_{j-1}|| + \frac{\lambda_1}{\lambda_1-\lambda_2}\sum_{k=1}^{j-1} ||\boldsymbol{\alpha}_k||\cdot||\boldsymbol{\alpha}_{j-1-k}|| \qquad \text{(C.9)}
\end{aligned}$$

Defining q_j as the upper bound of $||\boldsymbol{\alpha}_j||$, we have a new recurrence formula of q_j for $j > 2$ as

$$q_j = hq_{j-1} + g\sum_{k=0}^{j} q_k q_{j-k} \qquad \text{(C.10)}$$

with $q_0 = 0$, $q_1 = ||\boldsymbol{\alpha}_1||$ and $q_2 = ||\boldsymbol{\alpha}_2||$, where $h := \frac{1}{\lambda_1-\lambda_2}$ and $g := \frac{\lambda_1}{\lambda_1-\lambda_2}$. Further defining the generating function $G(z) = \sum_{j=0}^{\infty} q_j z^j$, we can derive that the generating function $G(z)$ follows

$$gG^2(z) = (1-hz)G(z) + (gq_1^2 + hq_1 - d_2)z^2 - q_1 z \qquad \text{(C.11)}$$

We define the functional equation $F(z,G) = gG^2(z) - (1-hz)G(x) - (gd_1^2 + hd_1 - d_2)z^2 + d_1 z$. Bender [223] shows that if there exist real positive numbers $r > 0$ such that the function equation $F(r,G(r)) = 0$ and $\frac{\partial F(z,G(z))}{\partial G}\big|_{z=r} = 0$, the convergence order of positive coefficients q_j follows $\lim_{j\to\infty}\sqrt[j]{q_j} = r^{-1}$. In our case, we can compute that $r = \frac{h+2gd_1-2\sqrt{gd_2}}{h^2+4hgd_1+4g^2d_1^2-4gd_2}$. According to Lemma 5.1, the radius of convergence for the prevalence expansion follows

$$R = \frac{1}{\lim_{j\to\infty}\sqrt[j]{|b_j|}} \geq \frac{1}{\lim_{j\to\infty}\sqrt[j]{||\boldsymbol{\alpha}_j||}} \geq \frac{1}{\lim_{j\to\infty}\sqrt[j]{q_j}} = r \qquad \text{(C.12)}$$

Thus, we can obtain a lower bound of the radius of convergence for the prevalence expansion

$$R_{lb} = \frac{h+2gd_1-2\sqrt{gd_2}}{h^2+4hgd_1+4g^2d_1^2-4gd_2} \qquad \text{(C.13)}$$

Figure C.1 shows the radius of convergence R and the proposed lower bound R_{lb} of (C.13) in more than 4.5×10^4 realizations in random sparse graphs. The radius of convergence R and the lower bound R_{lb} present similar behaviours, but the proposed R_{lb} only provides a loose lower bound for the radius of convergence R, which hints that the precise radius of convergence R cannot be inferred only by the eigenvalues of the adjacency matrix.

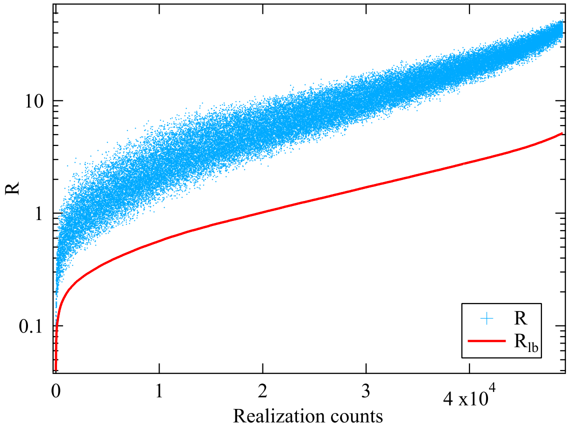
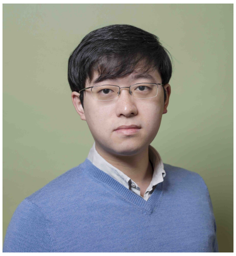


Figure C.1: Radius of convergence R for 4.5×10^4 realizations of connected ER random graphs sorted by the estimated lower bound R_{lb}. The red line represents the proposed lower bound of the radius of convergence R_{lb}. The ER random graph $G_p(N)$ is generated with the network size uniformly chosen in $N \in [10, 120]$ and the link density uniformly chosen in $p \in [0.05, \frac{2}{N} + 0.3]$.

CURRICULUM VITÆ



Zhidong He was born in Shanghai (China) on January 25th, 1989. He obtained his B.Sc. degree in School of Electronic Information and Electrical Engineering, as well as his B.A. degree in School of Foreign Languages, at Shanghai Jiao Tong University in 2011. He obtained his M.Sc. degree in College of Control Science and Engineering at Zhejiang University in 2014. He worked as a national certification engineer in Shanghai Institute of Quality Inspection and Technical Research after graduation. Since Oct 2015, He has been a Ph.D. researcher in the Networks Architectures and Services (NAS) group, faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, under the supervision of Prof. Piet Van Mieghem.

His research focuses are dynamics on networks, robustness assessment for networked systems, optimization algorithms and wireless networks.

LIST OF PUBLICATIONS

10. **Z. He** and P. Van Mieghem, *Heterogeneity induced localization of transport congestion on networks*, preparing.

9. **Z. He**, K. Navneet, W. van Dam and P. Van Mieghem, *Robustness assessment of multimodal freight transport network*, under submission.

8. **Z. He**, *Optimization of convergence rate via algebraic connectivity*, arXiv:1912.06536, under submission.

7. **Z. He** and P. Van Mieghem, *Prevalence expansion in NIMFA*, Physica A: Statistical Mechanics and its Applications, 2020, 540: 123220.

6. P. Sun, R. E. Kooij, **Z. He** and P. Van Mieghem, *Quantifying the robustness of network controllability*, 4th International Conference on System Reliability and Safety (ICSRS 2019), 20-22 November, Rome, Italy.

5. **Z. He**, P. Sun and P. Van Mieghem, *Topological approach to measure network recoverability*, 11th International Workshop on Resilient Networks Design and Modeling (RNDM 2019), 14-16 October, Nicosia, Cyprus. *(received the 2019 James P. G. Sterbenz best paper award.)*

4. **Z. He** and P. Van Mieghem, *Optimal induced spreading of SIS epidemics in networks*, IEEE Transactions on Control of Network Systems, 2018, 6(4): 1344-1353.

3. **Z. He** and P. Van Mieghem, *The fastest spreader in SIS epidemics on networks*, The European Physical Journal B, 2018, 91(5): 77.

2. **Z. He** and P. Van Mieghem, *The spreading time in SIS epidemics on networks*, Physica A: Statistical Mechanics and its Applications, 2018, 494: 317-330.

1. **Z. He**, *Extremal optimisation approach to joint routing and scheduling for industrial wireless networks*, International Journal of Ad Hoc and Ubiquitous Computing, 2016, 22(3): 164-173.