

Characterization and Design of Complex Networks

Characterization and Design of Complex Networks

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 20 oktober 2014 om 12:30 uur

door

Cong Li

bachelor telecommunicatie engineering
Jilin Universiteit, China
geboren te Harbin, Heilongjiang Provincie, China.

Dit proefschrift is goedgekeurd door de promotor:

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

Copromotor Dr. ir. H. Wang

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. P.F.A. Van Mieghem	Technische Universiteit Delft, promotor
Dr. ir. H. Wang	Technische Universiteit Delft, copromotor
Prof. dr. ir. R.E. Kooij	Technische Universiteit Delft
Prof. dr. F.H.J. Redig	Technische Universiteit Delft
Prof. dr. E. Estrada	University of Strathclyde
Prof. dr. C.J. Stam	Vrije Universiteit Amsterdam
Dr. D. Li	Beihang Univeristy
Prof. dr. ir. G. Jongbloed	Technische Universiteit Delft, reservelid



Alma mater: Technische Universiteit Delft

ISBN 978-94-6186-364-5

This research was supported by the China Scholarship Council.

Keywords: Complex Networks, Network Metrics, Spectral Metrics, Metric Correlations, Eigenvalues/Eigenvectors, Epidemic Threshold, Brain Networks, Opinion Model, Network Design

Copyright © 2014 by C. Li

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

Printed in the Netherlands.

Published by Uitgeverij BOXPress, 's-Hertogenbosch

Typeset by the author with the L^AT_EX Documentation System.

Author email: licong1986@gmail.com

To my husband and our parents

To the memory of youth

Summary

CAN you imagine how many networks you are on, when you check your Facebook or Twitter on a train? Your brain network guides your behavior and your metabolic network offers you energy. You and other passengers form a social network. Be careful, a virus may be spreading on the social network in your carriage! When you use your phone or laptop to access Facebook or Twitter, your devices are on Internet, your profiles are on the World Wide Web (WWW) and electricity is supplied by the electric power grid to your devices. Moreover, Facebook and Twitter are considered to be online social networks. Now, your phone is ringing, yes, you are on a communication network! Wait, wait... you are on a train, which is a part of a transportation network. Amazingly, complex networks have become a necessary part of our everyday life. This thesis opens with an introduction to the history of complex network studies and the essential primary knowledge, such as network metrics, network models and dynamic processes. Motivated by a better understanding of real networks, researchers explore the properties observed in real topologies. Plenty of metrics have been proposed to quantify the properties of networks, and further to characterize networks.

An essential question arises: “*How can we characterize our networks more efficiently and sufficiently?*”. For example, we can characterize a real network from different aspects: whether the network is well connected; whether the nodes in the network are near to each other; whether the neighbors of a node also tend to be neighbors; whether the nodes with similar degree are connected; whether the network is vulnerable to virus attacks, *etc.* If we use a number of metrics to characterize a network, is there any redundancy? However, if we only use one metric, is it sufficient to describe the network? Part I addresses the issues of selecting a representative set of metrics to efficiently and sufficiently characterize a network. The metric correlation pattern in studied functional brain networks is verified to be consistent with what we found in the Erdős-Rényi random graph model. Besides characterizing the whole network, quantifying the nodes in a network is also important. For example, when two companies compete for customers, they both try to prevail on influential customers to use their products. When they adopt different marketing strategies to select the influential customers, the competition results will be different. Which cus-

tomers should be selected as the “Very Important Persons (VIP)”: the customers with a lot of friends, the customers with a high reputation, the ones possessing impressionable friends or just random ones? Part I also studies the centrality metrics which are used to rank the importance of nodes. The correlations between centrality metrics are explored in network models and real-world networks. In addition, the centrality metrics are applied to an opinion dynamic process to help one opinion win the competition between two opinion groups. We show that the nodes in a network could be characterized efficiently by using a low-complexity metric to approximate a correlated high-complexity centrality metric.

A better understanding of the network characteristics could give rise to a better grasp on its dynamical and functional behavior. We compare two susceptible-infected-susceptible (SIS) mean-field approximations, with an ε -SIS model as the benchmark. The N-intertwined mean-field approximation (NIMFA) is shown to be a better approximation. A particular spectral metric, the spectral radius, is considered a better quantifier of the robustness to virus propagation in complex networks than the degree diversity. Part II mainly focuses on how we can make our networks less vulnerable to viruses. Minimizing the spectral radius could increase the epidemic threshold. Several approaches to enhance the epidemic threshold are presented in Part II. Spectral radius minimization by link or node removal is an NP-hard problem. The strategy that removes the links with the largest products of components of the principal eigenvector is shown to be superior to other strategies in most cases. Bounds for the decrease of the spectral radius by link or node removal are provided. Epidemics have so far been mostly studied in undirected networks. However, many real-world networks, such as online social networks and the WWW, on which information, emotion or malware spreads, are directed networks. Part II proposes two algorithms to generate directed networks with a given directionality. We declare that the epidemic threshold could be enlarged by raising the directionality of a directed network, via link rewiring or link resetting.

Finally, we conclude the main results, list the contributions of this thesis and suggest some research problems for further works.

Cong Li
Delft, April 2014

CONTENTS

1	Introduction	1
1.1	Why Studying Complex Networks	1
1.2	History of Complex Networks	2
1.3	Problem Statement	4
1.4	Thesis Outline	4
1.4.1	Part I: Network Characterization	4
1.4.2	Part II: Network Design	5
2	Complex Networks: Characteristics, Structures and Dynamics	7
2.1	Characteristics of Complex Networks: Graph Metrics	7
2.1.1	Structural Metrics	8
2.1.2	Spectral Metrics	10
2.1.3	Centrality Metrics	11
2.2	Structures of Complex Networks: Network Models	13
2.2.1	Bipartite Graphs	13
2.2.2	Regular Graphs	14
2.2.3	Erdős-Rényi Random Graphs	15
2.2.4	Watts-Strogatz Small-world Graphs	15
2.2.5	Bárabasi-Albert Graphs	16
2.2.6	Generalized Random Graphs	16
2.2.7	Directed Networks	17
2.3	Dynamics of Complex Networks: Processes on Networks	17
2.3.1	Epidemics in Networks	17
2.3.2	Opinion Dynamics	18
I	NETWORK CHARACTERIZATION	19
3	Correlations Between Structural and Spectral Metrics	21
3.1	Introduction	22

3.2	Analytic Relations Between Network Metrics	24
3.2.1	General Relations	24
3.2.2	Analytic Relations in Erdős-Rényi Random Graphs	26
3.2.3	Analytic Relations in Barabasi-Albert Graphs	27
3.3	Metric Correlations in Network Models	28
3.3.1	Erdős-Rényi Random Graphs	28
3.3.2	Barabasi-Albert Power-law Graphs	30
3.3.3	Watts-Strogatz Small-world Graphs	32
3.4	Metric Correlations in Functional Brain Networks	33
3.4.1	Metric Correlations in Unweighted Functional Brain Networks Transformed with Fixed Threshold T	33
3.4.2	Metric Correlations in Unweighted Functional Brain Networks Transformed with Fixed Link Density p	35
3.5	Chapter Conclusions	36
4	Correlations Between Centrality Metrics	39
4.1	Introduction	40
4.2	Analyzing the Correlation Between Degree and Principal Eigenvector	41
4.2.1	Properties of the Principal Eigenvector	42
4.2.2	Relation Between the Degree Vector and the Principal Eigenvector	44
4.2.3	Correlation Between the Principal Eigenvector and the Degree Vector	45
4.2.4	Application: Degree <i>vs.</i> Principal Eigenvector Strategy in Minimizing the Spectral Radius	46
4.2.5	Section Conclusions	47
4.3	Analyzing the Correlations Between Centrality Metrics	48
4.3.1	Pearson Correlation Coefficients Between Centrality Metrics	48
4.3.2	Centrality Similarities $M_{A,B}(\Upsilon)$ Between Centrality Metrics	48
4.4	Theoretical Analysis	51
4.4.1	Expectation and Variance of the Degree Masses	51
4.4.2	Correlation Between the Degree Masses and the Principal Eigenvector	52
4.5	Application to the Inflexible Contrarian Opinion (ICO) Model	53
4.5.1	The ICO Model	54
4.5.2	Strategies of Selecting Inflexible Contrarians Using Centrality Metrics	54
4.5.3	Comparison of Inflexible Contrarian Selection Strategies	54
4.6	Chapter Conclusions	57
5	Network Metrics for Epidemic Thresholds of SIS Approximations	59
5.1	Introduction	59
5.2	Description of the ε -SIS Model and the Mean-field Approximations	61
5.2.1	The ε -SIS Spreading Model	61
5.2.2	The Pastor-Satorras & Vespignani HMF Approximation	61
5.2.3	The N-intertwined Approximation	63
5.3	The Steady-state Infection in the Model and Two Approximations	64
5.3.1	The ε -SIS Spreading Model	64
5.3.2	The Pastor-Satorras & Vespignani HMF Approximation	65
5.3.3	The N-intertwined Approximation	66
5.3.4	Asymptotics for Large τ	66

5.4	Comparison of the Steady-state Fraction of Infected Nodes Versus τ . . .	67
5.4.1	Complete Bipartite Graphs	68
5.4.2	Star Graphs	68
5.4.3	Complete Graphs	69
5.4.4	Square Lattice Graphs	70
5.4.5	Path Graphs	70
5.4.6	Erdős-Rényi Random Graphs	71
5.4.7	Bárabasi-Albert Scale-free Graphs	71
5.4.8	Watts-Strogatz Small-world Graphs	72
5.5	Analytic Comparison of the Epidemic Thresholds $\tau_c^{(1)}$ and τ_c^{HMF}	73
5.6	Chapter Conclusions	74
 II NETWORK DESIGN		 75
6	Increasing the Epidemic Threshold by Link or Node Removals	77
6.1	Introduction	78
6.2	The <i>Spectral Radius Minimization Problem</i> is NP-hard	79
6.2.1	The Link Spectral Radius Minimization Problem is NP-hard	79
6.2.2	The Node Spectral Radius Minimization Problem is NP-hard	82
6.3	Removing the Links to Decrease the Spectral Radius	85
6.3.1	Spectral Graph Theory	85
6.3.2	Strategies to Minimize the Largest Eigenvalue by Link Removal	91
6.3.3	Scaling Law of $(\lambda_1(A) - \lambda_1(A_m))_{\text{optimal}}$	94
6.4	Removing the Nodes to Decrease the Spectral Radius	96
6.4.1	The Node Whose Deletion Reducing λ_1 Most	96
6.4.2	The Bounds for λ_1 When Nodes are Removed	98
6.5	Chapter Conclusions	103
7	Increasing the Epidemic Threshold by Link Rewiring or Resetting	105
7.1	Introduction	105
7.2	Algorithm Description	107
7.2.1	Degree-preserving Rewiring	108
7.2.2	In-degree and Out-degree Preserving Rewiring Algorithm (IOPRA)	108
7.2.3	Link Resetting Algorithm (LRA)	108
7.3	Spectral Properties in Directed Networks	109
7.3.1	Spectral Radius of Directed Networks	109
7.3.2	Principal Eigenvector in Directed Networks	113
7.3.3	Spectral Gap of Directed Networks	114
7.3.4	Algebraic Connectivity of Directed Networks	115
7.4	Effects of the Assortativity on λ_1 of Directed Networks	116
7.5	Chapter Conclusions	118
8	Conclusions and Future Work	119
8.1	Main Conclusions	119
8.2	Contributions Summary	122
8.3	Directions for Future Work	123

A Structural Metric Correlations in Network Models	125
B Introduction of Real-world Networks	129
C Pearson Correlation Coefficient Between Centrality Metrics	133
D Algorithms for Generating Directed Networks with Given Directionality	137
E Eigenvalues of the Directed Networks	139
List of Abbreviations	141
List of Symbols	143
Bibliography	145
Acknowledgements	159
Publications by the Author	161
Curriculum Vitae	163

CHAPTER 1

Introduction

“Everything existing in the universe is the fruit of chance and necessity.”

Democritus, c. 460 - c. 370 B.C.

1.1 Why Studying Complex Networks

THE whole universe is organized by all kinds of systems. Scientists are interested in what the systems are composed of, how the components in the systems are connected together, how the systems behave, *etc.* However, the systems are too complicated to be understood, since a component in a system might be a complex system. To make the problems simpler, researchers ignore the intricacy of individuals and consider a given system as a network, by presenting the components as the nodes and the connections as the links. The profound knowledge of complex networks will provide a strong foundation for understanding the systems, or even the universe. Scientists analyze, model and understand networks by using tools from wide range of research fields, such as mathematics, computations, statistics, or even medicine and engineering. The six degree of separation theory has revealed the small-world phenomenon in our life. The “rich get richer” generative model has been proposed as mechanism to explain the power-law degree distribution in real networks. The average shortest path length in the human functional brain network has been shown to be negatively correlated with IQ. The casualty of power outage caused by a cascading failure has inspired researchers to investigate how to strengthen the robustness of systems in our daily life. Although researches of complex networks are various, they all illuminate that the complex network theory is a useful tool to explore the unknown world. Therefore, if you are interested in a system, you can transform it into a network, then you will find out that a number of methods are available to help you analyze and understand your system.

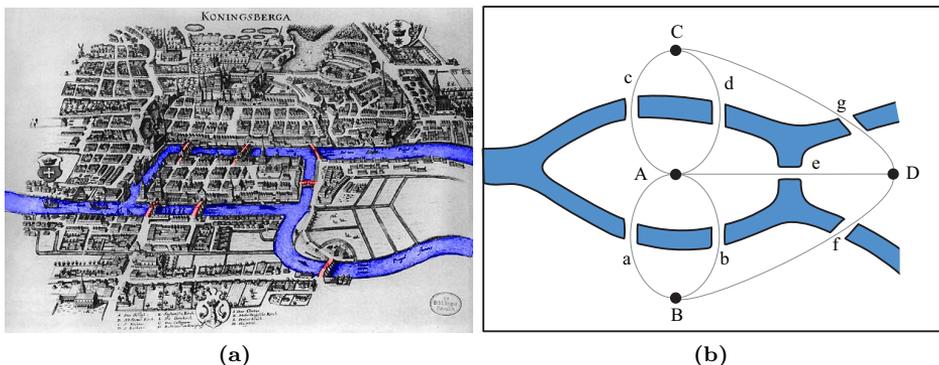


Figure 1.1: (a) Map of Königsberg’s in Euler’s time, with seven bridges joining the mainlands. The seven bridges of Königsberg problem was to find a walk through the city that would cross each of the seven red bridges once, and only once. Every bridge must be crossed completely every time. The starting point and the ending point can be any mainland. (b) Euler simplified the problem into a graph traversal problem and proved that the problem has **no** solution. To fulfill the requirement, every node in the graph, except possibly the starting and the ending nodes visited, should be connected by an even number of bridges. However, the numbers of links passing through the four nodes in (b) are all odd.

1.2 History of Complex Networks

From the communication networks to electric power grids, from the interconnected computer networks (Internet) to the World Wide Web (WWW), and from the social networks to the economic networks, complex networks have been pervading all of our daily life. The structure of complex networks has been mainly studied by mathematical graph theory. The basic idea is that a network consists of a set of items, which we call nodes, with connections between them, called links.

Graph theory was born in 1735, when Leonhard Euler published the solution of the Königsberg bridge problem, as illustrated in Fig 1.1. The historically notable problem shows the strength of graph theory, simplifying the complex real-world networks into a basic model. The study of graph theory usually focuses on the strong relation between the network structures and the network properties. In the 1960s, two Hungarian mathematicians, Paul Erdős and Alfréd Rényi, successfully combined the concepts of graph theory with probability theory, and established a new branch of graph theory—random graph theory [1, 2].

Meanwhile, besides the developments of mathematical graph theory, researchers had made breakthrough in unraveling the properties of social networks with experimental methods. The most famous one is the *small-world experiment* conducted by Stanley Milgram, which concluded that people in the United States are separated by six people on average in a 1969’s report [3]. The “six degrees of separation” suggests that human society has a small-world property. To verify the accuracy, some small communities, such as mathematicians and actors, had also done some small-world experiments. In the same year, a mathematician Casper Goffman defined the *Erdős number*¹, which describes the

¹Erdős number can be calculated on website: <http://www.oakland.edu/enp/compute/>

“collaborative distance” between a mathematician and Paul Erdős. The distribution of Erdős number is that, with a mean Erdős number of 4.65, almost everyone has an Erdős number of less than 8.

Table 1.1: *Kevin Bacon Number*

Kevin Bacon No.	# of People
0	1
1	2822
2	327146
3	1132453
4	283932
5	21784
6	2324
7	260
8	20

The *Kevin Bacon Number*² is an application of the same idea to the movie industry, connecting actors to Kevin Bacon in the smallest number of links possible. Two actors are linked together if they’ve been in one movie together. For instance, Jackie Chan and Gary Oldman both appeared in the movie “Kong Fu Panda 2 (2011)”, Gary Oldman and Kevin Bacon cooperated in the movie “Criminal Law (1988)”, but Jackie Chan and Kevin Bacon have never acted together, thus, Jackie Chan has a Kevin Bacon

number of 2. The average Kevin Bacon number is only 3.002. The following decades had witnessed the crucial conversion of the exploration of complex networks. The research interests were shifting from small networks to large-scale networks, and from regular and static networks to irregular, complex and dynamic real-world networks. The most striking motivation is the Internet [4] and WWW [5] use, which is an essential part of our life, increased rapidly from the middle of 1990s. Thanks to these changes, the approaches for network analysis had very fruitful improvements. The development of statistical methods and computer technology provided supports for analyzing the large-scale networks. At the end of twentieth century, two papers, that by Watts and Strogatz on small-world networks [6], appeared in Nature in 1998, and that by Barabási and Albert on scale-free networks [7], stood out as one of the notable milestones in the field of network theory. These two master pieces in network theory offer the possibility to study the properties of large databases of real-world networks.

Since the beginning of 2000s, thousands of papers and several popular scientific books [8, 9, 10, 11, 12, 13, 14] have appeared on the modern theory of complex networks. These books mainly focused on the following aspects: the structure and function of complex networks, the modeling of real-world networks, and various dynamical processes in networks. A large number of new measures have been developed to characterize network properties, including the clustering coefficient [15, 16, 13], the assortativity [17, 18], the spectral properties (such as the graph resistance [19]) and the centrality measures (such as the k-shell [20, 21], the leverage [22] and the degree masses [23]). These measures lead to a better understanding of the structure of complex networks, as well as give important insights into the dynamical and functional behaviors in networks. Moreover, the network science has been applied to broader fields. Statisticians apply the opinion model to predict the results of US presidential elections [24]. Neurologists create functional brain networks with brain magnetic activity records and use properties of brain networks to predict brain functioning such as cognitive performance [25, 26, 27, 28]. Epidemiologists are interested in the virus spreading processes in populations [29, 30, 31]. Recently, the study of complex networks has been expanded to networks of networks [32]. The topology [33], the spectral measures [34, 35] and the robustness [32, 36] of interdependent networks are studied.

²Kevin Bacon Number can be calculated on website: <http://www.cs.virginia.edu/oracle/>

1.3 Problem Statement

Researchers have proposed abundant various metrics to characterize networks from different perspectives. Previous studies of the network metrics contribute insights into the properties of complex networks. However, we can not take into account all properties of a real network when we characterize a network. We also have no idea which metrics are the most important and essential ones for characterizing a network. Thus, we are interested in how to select a set of representative metrics to characterize networks both efficiently and sufficiently. Moreover, characterizing the nodes in a network is also essential for understanding the structure of networks. How to characterize nodes efficiently is still an open problem. This thesis will explore the low-complexity metrics that can be used to approximate a high-complexity centrality metrics. Furthermore, the structure of a network affects the functions of the network, which are quantified by the metrics. Hence, we aim to design the structure of networks to make networks possess better functions (such as the ability of suppressing virus spreading). The goal of this thesis is to produce a framework to help address the mentioned issues.

1.4 Thesis Outline

The outline of the thesis is schematically illustrated in Figure 1.2. The thesis consists of eight chapters and is structured into two main parts: network characterization and network design.

- *Chapter 1* (current) first presents the reasons for studying complex networks and the history of complex networks. The problem statement for the research of this thesis is also given. Then, we introduce our framework to study the network characterization and design.
- *Chapter 2* is meant as an introductory chapter to complex networks. It exposes the necessary primary knowledge to understand complex network theory. The network metrics, network models and dynamic processes are introduced in this chapter. The metrics are classed into three taxonomies: the structural metrics, the spectral metrics and the centrality metrics. The models used in the thesis are introduced in detail. Moreover, two dynamic processes discussed in this thesis are also presented in this chapter.

1.4.1 Part I: Network Characterization

In this part, we study how to characterize a network with a given degree distribution both sufficiently and effectively, and which high-complexity centrality metric could be approximated by a strongly correlated low-complexity one. We also try to point out which approximation for the epidemic threshold is better.

- *Chapter 3* investigates the correlations between structural and spectral metrics. We present how to choose a representative set of metrics to characterize a network with a given degree distribution. The spectral metrics are shown to be essential for network characterization. We also verify the metric correlations in unweighted functional brain networks.

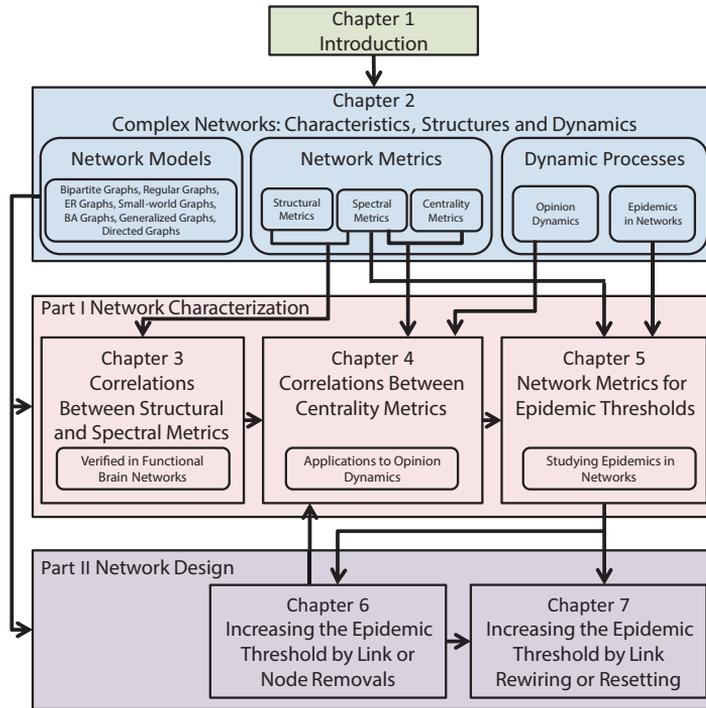


Figure 1.2: Schematic depiction of the present thesis.

- *Chapter 4* further explores the correlations between centrality metrics by the Pearson correlation coefficient and the centrality similarity. The properties of the principal eigenvector are studied. We find the conditions under which the degree vector and the principal eigenvector are strongly correlated. A novel centrality, the degree mass, is proposed in this chapter. The strong correlation between certain metrics are shown and theoretically proven. The centrality metrics are applied to study the opinion model.
- *Chapter 5* focuses on comparing two susceptible-infected-susceptible (SIS) mean-field approximations, with the ε -SIS spreading model as a benchmark, for different network types. We examine the epidemic threshold and the steady-state fraction of infected nodes of the mean-field approximations and the ε -SIS model in networks with different degree distributions.

1.4.2 Part II: Network Design

In this part, we focus on how to design a network with a high epidemic threshold.

- *Chapter 6* studies how to alter a network to enlarge the epidemic threshold of the network. The minimization of the spectral radius by removing m links (or nodes) is shown to be an NP-complete problem. The strategy that removes that link $l = i \sim j$ with largest product $(x_1)_i(x_1)_j$ of the components of the principal eigenvector x_1 belonging to the largest adjacency eigenvalue is presented to be superior to other strategies in most cases. The bounds for the decrease of the spectral radius by link (or node) removals are provided.
- *Chapter 7* introduces two algorithms to generate directed network with a given directionality ξ . The effects of ξ on spectral properties are studied. Important findings are that the spectral radius λ_1 decreases with the increase of the directionality ξ . Thus, we could use the rewiring or resetting methods to increase the directionality, equivalently, to enlarge the epidemic threshold.

Finally, *Chapter 8* concludes the present thesis, highlights the main contributions and provides the directions for future work.

CHAPTER 2

Complex Networks: Characteristics, Structures and Dynamics

*“Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less.”*

Marie Curie, 1867 - 1934

TRADITIONALLY complex networks have been studied in the form of mathematical graph theory and statical mechanics. As the developments in complex networks have taken place in areas, such as social networks, the dynamical evolution of network topology has also been researched extensively. This chapter first introduces the essential aspects that describe a network: the graph metrics and the network models. Next we mainly elaborate the dynamic processes on networks that will be investigated through this thesis.

2.1 Characteristics of Complex Networks: Graph Metrics

A complex network is represented as a graph $G = (\mathcal{N}, \mathcal{L})$, which consists of a set of nodes \mathcal{N} interconnected by a set of links \mathcal{L} . The number of nodes is denoted by $N = |\mathcal{N}|$ and the number of links is represented by $L = |\mathcal{L}|$. When nodes i and j are linked, i is said to be a neighbor of j . The number of neighbors of a node j is called its degree d_j . The basic law for the degree is $\sum_{j=1}^N d_j = 2L$. The degree of an arbitrary node is denoted by D . The graph G can be represented by an $N \times N$ adjacency matrix A , consisting of elements a_{ij} that are either one or zero depending on whether there is a link between nodes i and j . The adjacency matrix of an undirected graph is symmetry, on the contrary, a directed graph owns an asymmetric adjacency matrix (see Figure 2.1). The graphs mentioned in this thesis are simple, unweighted, without self-loops nor multiple links. The Laplacian matrix of G is an $N \times N$ matrix $Q = \Delta - A$, where $\Delta = \text{diag}(d_i)$.

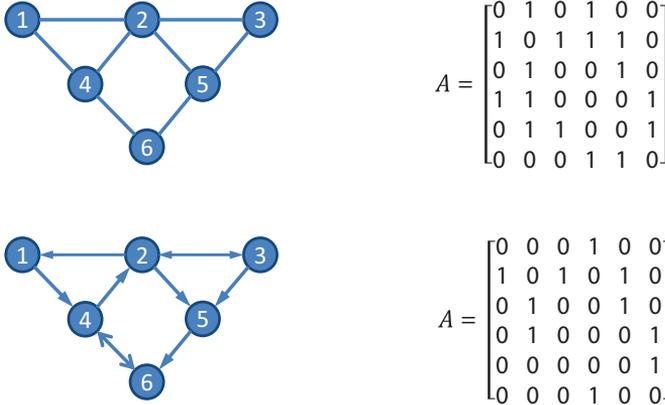


Figure 2.1: Examples of undirected and directed graphs and their adjacency matrices.

In this section, we introduce the network metrics that are widely studied in the literature, from the classical structural metrics to the spectral metrics (eigenvalue related metrics), and to the centrality metrics.

2.1.1 Structural Metrics

- Degree diversity κ

The degree diversity κ is defined [11] as

$$\kappa = \frac{E[D^2]}{E[D]} = \frac{Var[D] + E[D]^2}{E[D]}, \tag{2.1}$$

where $E[D]$ and $Var[D]$ are the mean and variance of D . Chung *et al.* [37] found that the degree diversity approximates the largest adjacency eigenvalue λ_1 in Erdős-Rényi random graphs, if $\kappa > \sqrt{d_{max}} \ln N$ (d_{max} is the maximum degree). Scale-free networks, where $Var[D] \rightarrow \infty$ as $N \rightarrow \infty$, are characterized by $\kappa \rightarrow \infty$, whereas regular networks, where $Var[D] = 0$, have $\kappa = E[D]$. Properties of dynamic processes on networks, such as the synchronization threshold in the mean-field theory of coupled oscillators in networks [38], the network percolation [13] and the epidemic thresholds [39], have all been stated to be related to $\kappa = \frac{E[D^2]}{E[D]}$, approximately.

- Assortativity ρ_D

“Mixing” in complex networks [18] refers to the tendency of network nodes to connect preferentially to other nodes with either similar or opposite properties. The mixing of the degree is computed via the degree correlation of connected node pairs, called assortativity [40],

$$\rho_D = 1 - \frac{\sum_{i \sim j} (d_i - d_j)^2}{\sum_{i=1}^N d_i^3 - \frac{1}{2L} \left(\sum_{i=1}^N d_i^2 \right)^2}. \tag{2.2}$$

Networks where high degree nodes preferentially connect to other high degree nodes, are assortative in the degree correlation ($\rho_D > 0$), whereas networks where high degree nodes connect to low-degree nodes, are disassortative ($\rho_D < 0$). Van Mieghem *et al.* [40, 14] have reformulated the assortativity as follows

$$\rho_D = \frac{N_1 N_3 - N_2^2}{N_1 \sum_{i=1}^N d_i^3 - N_2^2}, \quad (2.3)$$

where $N_k = u^T A^k u$ is the total number of walks with k hops. Newman [17] found that technological and biological networks are disassortative while social networks are assortative. The functional brain networks determined from EEG have also been found to be assortative [41].

- Clustering coefficient C_G

Two different clustering coefficients are frequently used. The first definition of clustering coefficient [13, 15, 16] C_G of a graph is the average clustering coefficient of all nodes, given as

$$C_G = \frac{1}{N} \sum_{v \in \mathcal{N} - \mathcal{N}^{(1)} - \mathcal{N}^{(0)}} c_G(v), \quad (2.4)$$

where \mathcal{N} is the set of all nodes and $\mathcal{N}^{(k)}$ is the set of nodes with degree k . The clustering coefficient of a node $c_G(v)$ characterizes the density of connections in the environment of a node v and is defined as the ratio of the number of links y connecting the $d_v > 1$ neighbors of v over the total possible $\frac{d_v(d_v-1)}{2}$, thus $c_G(v) = \frac{2y}{d_v(d_v-1)}$. Note that $c_G(v) = 0$, for a node v with degree 0 or 1.

The second one is based on the following definition for undirected unweighted networks [42]

$$C_G = \frac{3\blacktriangle_G}{N_\Delta}, \quad (2.5)$$

where \blacktriangle_G is the number of triangles in the network and $N_\Delta = \sum_{i=1}^N \binom{d_i}{2}$ is the number of connected triples.

The difference between the two definitions is that Eq. (2.4) is the average of the connection density among the neighbors of each node, while Eq. (2.5) is the average of the probability that a triangle is formed upon each triple in the network. In this thesis, we consider the effect of degree distribution, so we use the definition Eq. (2.4) to calculate the clustering coefficient to each node.

- Average hopcount $E[H]$ and Global efficiency $E[\frac{1}{H}]$

The hopcount H_{ij} is the number of links or hops in the shortest path between node i and node j . The maximal hopcount H_{\max} among all node pairs is the diameter of a graph. If the average hopcount of a network approximates that of the corresponding Erdős-Rényi random graph with same number N of nodes and link density p ($E[H_G] \approx E[H_{G_p(N)}]$), and the clustering coefficient always $C_G > C_{G_p(N)}$, then, the network possesses the small-world property.

When a network is disconnected, the shortest paths between some node pairs are infinite, so the average hopcount of the network can not be computed. In this situation,

we compute the average reciprocal hopcount $E[\frac{1}{H}]$, which is called the global efficiency and widely studied in neuroscience [15, 43]. In addition to the global efficiency in [15], a local efficiency is defined as $E_{loc} = 1/N \sum_{i \subseteq G} E(G_i)$, where G_i is the sub-graph of the neighbors of i . The local efficiency plays a role similar to the first definition of clustering coefficient C_G .

2.1.2 Spectral Metrics

- Spectral radius (the largest adjacency eigenvalue) λ_1

We denote the set of eigenvalues of the adjacency matrix A as $\lambda_N \leq \lambda_{N-1} \leq \dots \leq \lambda_1$, where the largest eigenvalue λ_1 is called the spectral radius. The eigenvalues of the adjacency matrix are real [15]. The largest eigenvalue λ_1 is a powerful character of dynamic processes on networks such as virus spreading and synchronization processes [38]. The inverse of the largest eigenvalue λ_1 characterizes the threshold of the phase transition, which specifies the onset of a remaining fraction of infected nodes and of locked oscillators respectively, of both virus spreading [31] and synchronization of coupled oscillators [44] in networks. Restrepo *et al.* [38] discovered that λ_1 can be approximated by N_3/N_2 , where N_k is the total numbers of walks with k hops. Recently, Van Mieghem *et al.* [40] proved that N_3/N_2 is a lower bound of the largest adjacency eigenvalue λ_1 .

- Effective graph resistance R_G

The effective graph resistance (which is also called the Kirchhoff index) originated from the field of electric circuit analysis [19, 45]. Assuming a network as an electrical circuit where the resistance of each link is 1, the effective graph resistance is defined as the accumulated effective resistance between all pairs of nodes. It measures the ease of communication in a graph [19, 46]. The equivalent spectral expression for the effective graph resistance is [14]

$$R_G = N \sum_{k=1}^{N-1} \frac{1}{\mu_k}, \quad (2.6)$$

where μ_k is the k -th largest eigenvalue of the Laplacian matrix Q .

- Algebraic connectivity μ_{N-1}

The eigenvalues of the Laplacian matrix Q are ordered as $0 = \mu_N \leq \mu_{N-1} \leq \dots \leq \mu_1$, and $\mu_{N-1} > 0$ if and only if the graph G is connected. The second smallest eigenvalue μ_{N-1} of Q is called the algebraic connectivity. It was first studied by Fiedler [47]. A large value of algebraic connectivity characterizes strong network robustness regarding to e.g. a) the difficulty to cut the network into separated sub-parts [48] and b) enhanced synchronizability and fast convergence [49, 50].

- Ratio μ_1/μ_{N-1}

The ratio of the largest eigenvalue μ_1 and the second smallest eigenvalue μ_{N-1} of Laplacian is often claimed as an index of synchronizability of a graph [28]. The synchronizability mainly indicates whether the synchronized state of a dynamic on a graph will be stable

for a sufficiently large range of the parameters of the dynamic process [51, 52]. The larger the ratio is, the more difficult it is to synchronize the oscillators and vice versa [51]. The ratio is also referred to as the “paradox of heterogeneity”. It shows that (unweighted, undirected) networks with a more homogenous degree distribution synchronize more easily than networks with a more heterogeneous degree distribution [28]. In [14], Van Mieghem has explained that $\mu_1 > D_{\max}$ and $\mu_{N-1} \leq D_{\min}$. Hence, $\frac{\mu_1}{\mu_{N-1}} > \frac{D_{\max}}{D_{\min}}$ implying that the ratio μ_1/μ_{N-1} is larger for heterogeneous networks because D_{\max}/D_{\min} is larger, while homogeneous networks have smaller ratios μ_1/μ_{N-1} .

2.1.3 Centrality Metrics

- Degree mass $D^{(m)}$

The degree of a node i in a network G is the number of its direct neighbors,

$$d_i = \sum_{j=1}^N a_{ij} = (Au)_i,$$

where $u = (1, 1, \dots, 1)^T$ is the all-one vector. Here we propose a new set of centrality metrics, the degree mass, which is a variant of degree centrality. The m th-order degree mass of a node i is defined as the sum of the weighted degree of its m -hop neighborhood¹,

$$d_i^{(m)} = \sum_{k=1}^{m+1} (A^k u)_i = \sum_{j=1}^N \left(\sum_{k=0}^m A^k \right)_{ij} d_j,$$

where $m \geq 0$. The weight of the degree d_j is the number of walks² of length no longer than m from node i to node j . The weight of d_j is larger than the weight of d_l when node l is farther than node j from node i . The m th-order degree mass vector is defined $d^{(m)} = [d_1^{(m)}, d_2^{(m)}, \dots, d_N^{(m)}]$. The 0th-order degree mass is the degree centrality. The 1st-order degree mass of node i is the sum of the degree of node i and the degree of its nearest neighbors. When m is large, the m th-order degree mass is proportional to the principal eigenvector.

- Principal eigenvector x_1

The largest eigenvalue of the adjacency matrix A is λ_1 , also called the spectral radius [14]. The principal eigenvector x_1 corresponding to the spectral radius λ_1 satisfies the eigenvalue equation

$$Ax_1 = \lambda_1 x_1.$$

The j -th component of the principal eigenvector is denoted by $(x_1)_j$. The X_1 is the element in the principal eigenvector corresponding to a random node.

- Betweenness B_n

¹The m -hop neighborhood of a node i includes the node i and all nodes no further away than m hops from i .

²A walk from i to j is any sequence of edges that allows back and forth movement and repeated visits to the same node.

Betweenness was introduced independently by Anthonisse [53] in 1971 and Freeman [54] in 1977. The betweenness of a node i is the number of shortest paths between all possible pairs of nodes in the network that traverse the node

$$b_{ni} = \sum_{s \neq i \neq d \in \mathcal{N}} \frac{\sigma_{sd}(i)}{\sigma_{sd}},$$

where $\sigma_{sd}(i)$ is the number of shortest paths that pass through node i from node s to node d , and σ_{sd} is the total number of shortest paths from node s to node d . The betweenness B_n incorporates global information and is a simplified quantity for assessing the traffic carried by a node. Assuming that a unit packet is transmitted between each node pair, the betweenness b_{ni} is the total number of packets passing through node i [55].

- Closeness C_n

The closeness [56] of a node i is the average hopcount of the shortest paths from node i to all other nodes. It measures how close a node is to all the others. The most commonly used definition is the reciprocal of the total hopcount:

$$c_{ni} = \frac{N - 1}{\sum_{j \in \mathcal{N} \setminus \{i\}} H_{ij}},$$

where H_{ij} is the hopcount of the shortest path between nodes i and j , and $\sum_{j \in \mathcal{N} \setminus \{i\}} H_{ij}$ is the sum of the hopcount of the shortest paths from node i to all other nodes. Closeness has been used to identify central metabolites in metabolic networks [57].

- K -shell index K_s

The k -shell decomposition of a network allows us to identify the core and the periphery of the network. The k -shell decomposition procedure is as follows:

- (1) Remove all nodes of degree $d = 1$ and also their links. This may reduce the degree of other nodes to 1.
- (2) Remove nodes whose degree has been reduced to 1 and their links until all of the remaining nodes have a degree $d > 1$. All of the removed nodes and the links between them constitute the k -shell with an index $k_s = 1$.
- (3) Remove nodes with degree $d = 2$ and their links in the remaining networks until all of the remaining nodes have a degree $d > 2$. The newly removed nodes and the links between them constitute the k -shell with an index $k_s = 2$, and subsequently for higher values of k_s .

The k -shell is a variant of the k -core [58, 59], which is the largest sub-graph with minimum degree of at least k . A k -core includes all k -shells with an index of $k_s = 0, 1, 2, \dots, k$. An $O(m)$ algorithm for k -shell network decomposition was proposed in Ref. [20]. The k -shell index of the original infected node is a better predictor of the infected population in the susceptible-infectious-recovered (SIR) epidemic spreading process than other centrality metrics, such as the degree [21].

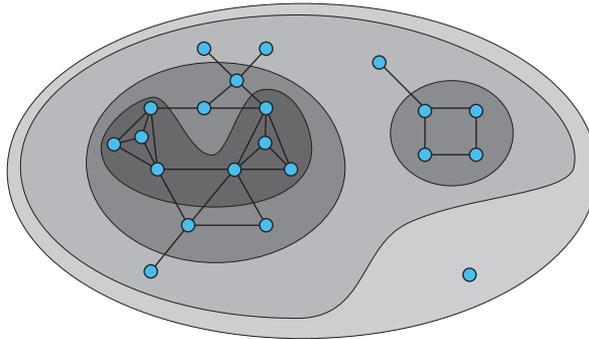


Figure 2.2: 0, 1, 2 and 3 shells (and cores) of a sample graph [20].

- Leverage L_n

Joyce *et al.* [22] introduced leverage centrality in order to identify neighborhood hubs in functional brain networks. The leverage measures the extent of the connectivity of a node relative to the connectivity of its nearest neighbors. The leverage of a node i is defined as

$$l_{ni} = \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} \frac{d_i - d_j}{d_i + d_j},$$

where \mathcal{N}_i is the directly connected neighbors of the node i . With the definition of l_{ni} and the range $[1, N - 1]$ of the degree d_i in connected networks, the leverage of a node i is bounded by $-1 + \frac{2d_i}{d_i + (N-1)} \leq l_{ni} \leq 1 - \frac{2}{d_i + 1}$. Hence the range of the leverage l_{ni} is $[-1 + 2/N, 1 - 2/N]$ and the equality occurs in star graphs and complete graphs K_N . The leverage of a node is high when it has more connections than its direct neighbors. Thus a high-degree node with high-degree nearest neighbors will probably have a low leverage.

2.2 Structures of Complex Networks: Network Models

Network anatomy is such important to characterize, because the structure of networks always affects the functions [60]. Numerous models of network structures have been proposed for studying the topological properties of real-world networks. This section presents the network models which will be frequently used throughout this thesis.

2.2.1 Bipartite Graphs

Bipartite graph is a graph whose nodes can be divided into two disjoint sets S_1 and S_2 , and each link connects a node in set S_1 and one in S_2 . However, the nodes within a set S_1 (or S_2) do not connect. Here we introduce some types of bipartite graphs. All graphs in Figure 2.3 are bipartite graphs, when we consider the circle nodes are in one set and the disc nodes are in the other set.

- Complete bipartite graph

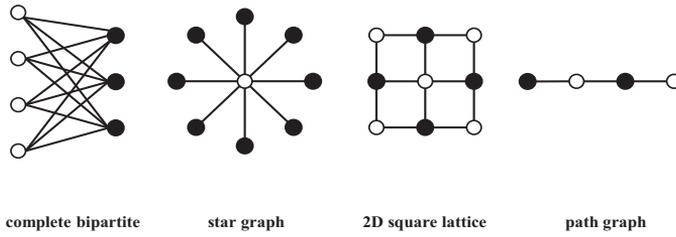


Figure 2.3: *Examples of bipartite graphs.*

A complete bipartite graph K_{M_1, M_2} consists of two disjoint sets S_1 and S_2 containing respectively M_1 and M_2 nodes. All nodes in S_1 are connected to all nodes in S_2 , while nodes within a set do not connect.

- Star graph

The star graph $K_{1, N-1}$ is a special complete bipartite graph where one of the disjoint sets contains only one node while the other set contains the rest of the nodes. The star graph is the basic computer network model.

- Lattice graph

The square lattice graph is a two-dimensional grid. Ignoring the boundary nodes, the square lattice can be regarded as a regular graph, where all nodes have the same degree ($d_i = 4$). The lattice is the basic model of a transport network (Manhattan grid) and is crucial in percolation theory [61, 62, 63]. Moreover, it is frequently used to study the network traffic [64].

- Path graph

The path graph is an example of a tree graph, in which every root node has only one branch and only the last root node is not branched at all.

2.2.2 Regular Graphs

A regular graph is a graph with all nodes having the same number of neighbors, *i.e.* each node j has the same degree $d_j = r$. The spectral radius $\lambda_1 = r$, in regular graphs, where $r = E[D] = D_{max}$. The differences $\lambda_1 - E[D]$ and $d_{max} - \lambda_1$ can be considered as measures for the irregularity of a graph [14]. Before modeling real-world networks by coupling dynamical systems with networks, the connection topology was assumed to be either completely regular or completely random [6].

- Complete graph

The complete graph K_N is a graph in which every node pair is connected. For a complete graph with N nodes, the degree of each node is $r = N - 1$.

- Ring lattice graph

The ring lattice graph is a graph with N nodes and each node connects to its r nearest neighbors by undirected links. The small-world network [6] was proposed basing on the ring lattice graph.

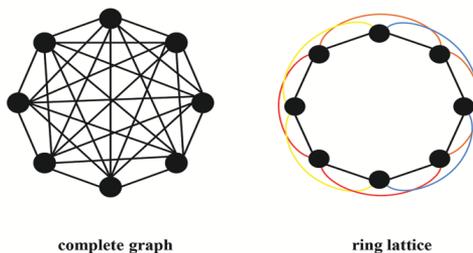


Figure 2.4: *Examples of regular graphs.*

2.2.3 Erdős-Rényi Random Graphs

The Erdős-Rényi (ER) random graph was first proposed by Paul Erdős and Alfréd Rényi in 1959 and 1960 [1, 2]. Thanks to their contributions, the Erdős-Rényi random graph $G_p(N)$ has become a frequently occurring random complex network model, where N is the number of nodes and p is the link density between any two nodes. The beauty of this model lies mainly in its mathematical simplicity—almost everything about its structure can be calculated analytically. The ER graphs are characterized by a binomial degree distribution with

$$\text{Prob}[D = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (2.7)$$

The average degree of nodes is $E[D] = p(N-1)$, in ER graphs, and the clustering coefficient is $C_G = p$. When p is small, the distribution converges on a Poisson distribution. Thus, in the sparse ER graphs, the degree distribution is $\text{Prob}[D = k] = e^{-E[D]} E[D]^k / k!$. Moreover, an ER random graph is connected, if $p > p_c \approx \frac{\ln N}{N}$ for large N , where p_c is the disconnectivity threshold.

2.2.4 Watts-Strogatz Small-world Graphs

A small-world graph refers to properties: (1) the average short path $E[H]$ is small, like that in Erdős-Rényi random graph; (2) the clustering coefficient C_G is high, like that in ring lattice. The structural properties of small-world networks have also been found in real-world networks, including social networks [65], neural networks [66] and biological oscillators [67]. The small-world graphs of Watts and Strogatz [6] can be generated from a ring lattice with N nodes and k links per node, by rewiring each link at random with probability p_r (see Figure 2.5). We choose the rewiring probability $p_r = 0.01$ in this thesis to generate graphs that have small-world properties, because that: (1) in small-world rewiring, $C(p_r)$ is the clustering coefficient of the small-world graph with rewiring probability p_r and $C(0)$ is the clustering coefficient of the ring lattice without rewiring. The dependence of the clustering coefficient ratio $C(p_r)/C(0)$ on N is very small, and $C(p_r)/C(0)$ decreases with the fraction p_r , as $C(p_r)/C(0) \approx (1-p_r^3)$ [68]; (2) the average shortest path $E[H]$ is much small even when p_r is small, as long as the size N of network is large enough [68].

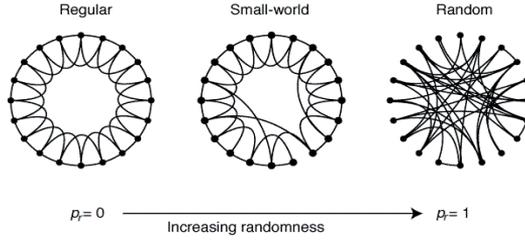


Figure 2.5: A ring lattice graph has its links rewired with probability p_r . Three realizations of the small-world network generating process are shown, for different values of p_r . For $p_r=0$, the network is regular, with large clustering coefficient C_G and average shortest path $E[H]$; for $p_r=1$, the network is random, with small clustering coefficient C_G and average shortest path $E[H]$. For an intermediate value of p_r , the graph is a small-world network: highly clustered like a regular graph, yet with a small average shortest path $E[H]$ like a random graph. The Figure is taken from [6].

2.2.5 Barabasi-Albert Graphs

The Barabasi-Albert (BA) graph is one of the most studied network models since, when N is large, it possesses a power-law degree distribution $\text{Prob}[D = k] \sim k^{-\alpha}$, which is a common property of many real-world networks, such as World Wide Web (WWW) and citation patterns in science [7]. The BA graph is generated basing on two key rules: growth and preferential attachment as follows. Starting with a small number m_0 of nodes, at each step a new node with $m (\leq m_0)$ links is added to the network. The m links are connected to the nodes that have already been in the system. The probability Π that a link will connect to an existing node i is linearly proportional to the degree of the node i :

$$\Pi = d_i / \sum_j d_j \quad (2.8)$$

The high degree nodes have a higher probability to be connected to. This is known as a “the rich get richer” phenomenon. Because every new node adds m links into the network, the average degree of BA graphs is $E[D] = 2m$ when N is large. The BA graph [7, 69] owns a power-law degree distribution with an exponent $\alpha = 3$.

2.2.6 Generalized Random Graphs

As the degree of a node is one of the most fundamental network characterizations, it is interesting to generate networks with a given degree distribution. The *configuration model* allows to built up random graphs with a given degree distribution. The probability of any two nodes being connected does not depend on the degrees of the nodes. The model is defined as follows. Given a degree distribution $\text{Prob}[D = k]$, a degree sequence $k = \{k_1, k_2, \dots, k_N\}$ can be selected, by comparing N random numbers $r(i) \in (0, 1]$ (where $i = 1, 2, \dots, N$) with the uniformed degree cumulative distribution function (CDF) $\text{Prob}[D \leq k]$. For example, if $\text{Prob}[D \leq 3] = 0.2$, $\text{Prob}[D \leq 4] = 0.3$ and the random number $r(i) = 0.23$, the degree k_i is chosen as 4. We can write the index i of each node k_i times in a vector v of length $2L$, when we consider that each node i is associated a number k_i of “stubs” (ends of links from a node). Then, the elements in vector v are randomly paired, correspondingly, we obtain a random matching of the stubs. The stub-matching process

is shown in Figure 2.6. The scale-free (SF) networks used in this thesis are generated by the *configuration model*.

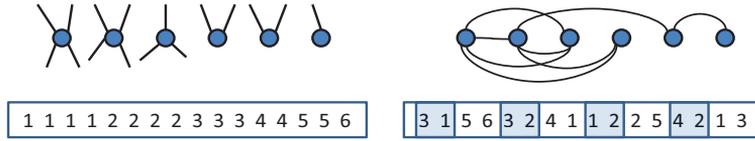


Figure 2.6: An example of the stub-matching process for building up a network by the configuration model.

2.2.7 Directed Networks

In a network, if node i is connected to node j (denoted by $i \rightarrow j$) then j is also linked to i (denoted by $j \rightarrow i$), one bidirectional link exists between nodes i and j ; and if either $i \rightarrow j$ or $j \rightarrow i$ exists, but not both in between the node pair i and j , a unidirectional link exists. The networks containing at least one unidirectional link are directed networks, whose adjacency matrix is asymmetric. Each node in a directed graph has a in-degree d_{in} and a out-degree d_{out} . The in-degree and out-degree of a node i are represented as $d_{in}(i) = \sum_j \mathbf{1}_{j \rightarrow i}$ and $d_{out}(i) = \sum_j \mathbf{1}_{i \rightarrow j}$. The directionality [70] of a directed network is defined as $\xi = L_{unidirectional}/L_{arcs}$, where the number of arcs (the number of nonzero elements in the adjacency matrix) $L_{arcs} = \sum_i \sum_j a_{ij} = u^T A u$, (u is the all-one vector), can also be calculated by $L_{arcs} = L_{unidirectional} + 2L_{bidirectional}$. A directed network with directionality ξ is denoted by $G^{(\xi)}$. The network $G^{(\xi)}$ reduces to a bidirectional network or an undirected network, when $\xi = 0$. The network $G^{(\xi=1)}$ is a directed network without any bidirectional link, when $\xi = 1$. Many important networks, such as the WWW and the metabolic networks, are directed networks.

2.3 Dynamics of Complex Networks: Processes on Networks

The functions of a network are usually expressed through the dynamic processes on the network. The percolation [71], synchronization [69], epidemics [69, 72, 73, 31], opinion dynamics [74, 75, 76] and the cascading failures [32, 77, 78] are the most widely studied dynamic processes. This section will only introduce the dynamic processes that will be studied in this thesis.

2.3.1 Epidemics in Networks

In epidemiology, modeling has been used in arranging, implementing and evaluating various prevention, therapy and control programs [79, 72]. Here we introduce two basic virus spreading models: the susceptible-infected-removed (SIR) model and the susceptible-infected-susceptible (SIS) model.

- SIR model

The SIR model assumes that each individual can be in one of three possible states, susceptible, infected and removed. Susceptible individuals are healthy persons and can

be infected by unhealthy individuals. Each infected individual can cure and become a removed individual. A removed individual can not be infected anymore, since it is immune to the virus.

- SIS model

In the SIS model, each node only has two states, healthy and infected. Healthy individuals are susceptible to the disease and can be infected. The infected individuals can cure and become healthy, but again susceptible after recovering from the disease.

The SIS epidemic process mentioned in this work is a simple continuous-time model for the spreading of a virus in a network, which belongs to the SIS model. The SIS epidemic process in network is described as follows. The arrival of an infection over a link and the curing of an infected node are assumed to be independent Poisson processes with rates β and δ , respectively. Only infected nodes can infect their healthy direct neighbors. The effective spreading rate is defined as $\tau = \frac{\beta}{\delta}$. The viral state of a node i at time t is specified by a Bernoulli random variable $X_i(t) \in \{0, 1\}$, where $X_i(t) = 0$ refers to a healthy node and $X_i(t) = 1$ to an infected node. Every node i at time t is either infected, with probability $v_i(t) = \text{Prob}[X_i(t) = 1]$ or healthy (but susceptible) with probability $1 - v_i(t)$.

2.3.2 Opinion Dynamics

There are a lot of opinion competitions in our life. For example, the Apple company competes with Microsoft, the McDonald's competes with Burger King and the two leaders competition in election. Various versions of the opinion model have been proposed [74], among which are the Sznajd model [80], the voter model [81] and the majority rule model [82]. The steady state of these opinion models is either consensus of a single opinion or equal concentrations of the two opinions [81, 82]. In real life, however, a stable coexistence with unequal concentrations of two opinions is commonly seen [75]. We will introduce two newly proposed opinion models [75, 76], in which stable coexistence of minority and majority opinions occurs.

- Nonconsensus opinion (NCO) model

Two opinions are randomly assigned to all agents (nodes): an agent will be assigned opinion A with a probability f and opinion B with a probability $1 - f$, when time $t = 0$. Then, at each time step, each agent adopts the opinion of the majority in its nearest neighbors and itself. In the case of a tie, an agent does not change its opinion. All of the updates are made simultaneously in parallel at each step. The system will reach a stable state, where the opinions A and B stably coexist, when f is above a critical threshold f_c .

- Inflexible contrarian opinion (ICO) model

The ICO model is a variant of the nonconsensus opinion (NCO) model. The initial state of the ICO model is the steady state of the NCO model. The ICO model further selects a fraction p_o of agents with opinion A as the inflexible contrarians, which will hold opinion B and never change their opinion again, but can influence the opinion of the other agents. Then, the two opinions compete with each other by the update rules of the NCO model. The system will reach a new stable state by following the above opinion dynamics rules.

Part I

NETWORK CHARACTERIZATION

CHAPTER 3

Correlations Between Structural and Spectral Metrics

“Everything should be made as simple as possible, but not simpler.”

Albert Einstein, 1879 - 1955

AN increasing number of network metrics have been applied in network analysis. If metric relations would be known better, we could more effectively characterize networks by a small set of metrics to discover the association between network properties/metrics and network functioning. In this chapter, we investigate the Pearson correlation coefficients between widely studied structural and spectral network metrics in three network models (Bárabasi-Albert graphs, Erdős-Rényi random graphs and Watts-Strogatz small-world graphs) as well as in functional brain networks of healthy subjects. The metric correlations, that we observed and theoretically explained, motivate us to propose a small representative set of metrics by including only one metric from each sub-set of mutually strongly dependent metrics. The following contributions are considered important: a) A network with a given degree distribution can indeed be characterized by a small representative set of metrics. b) Unweighted networks, which are obtained from weighted functional brain networks with a fixed threshold, and Erdős-Rényi random graphs follow a similar degree distribution. Moreover, their metric correlations and the resultant representative metrics are as well similar. This verifies the influence of degree distribution on metric correlations. c) Most metric correlations can be explained analytically. d) Interestingly, the most studied metrics so far, the average shortest path length and the clustering coefficient, are strongly correlated, thus, redundant. Whereas spectral metrics, though only studied recently in the context of complex networks seem to be essential in network characterizations. This representative set of metrics tends to both sufficiently and effectively characterize networks with a given degree distribution. In the study of

a specific network, however, we have to at least consider the representative set so that important network properties will not be neglected.

3.1 Introduction

After about a decade of extensive research on complex networks, numerous metrics have been introduced to quantify different features of complex networks [42, 27]. The computational complexity of network metrics can be high. Actually, network metrics can be strongly correlated in a certain type of graphs such as power-law graphs, indicating redundancy among them. On the other hand, each metric only partially captures the properties of a network. It would be helpful if it is possible to define a small representative set of network metrics that effectively characterize a given type of networks. Understanding the relations between network metrics is essential, in general, for complex network studies. In this chapter, we take neuroscience as an illustration.

Network science has recently been applied to neuroscience to understand the effect of the network structure on its functioning. The average shortest path length in the human functional brain network was shown to be negatively correlated with IQ [83, 84]. The small-world pattern and modularity tend to disappear in the brain networks of patient groups with e.g. brain tumors, epilepsy or Alzheimer [85, 86]. Well-studied metrics like degree diversity (κ), assortativity (ρ_D), clustering coefficient (C_G), average hopcount ($E[H]$), global efficiency ($E[\frac{1}{H}]$), spectral radius (λ_1), effective graph resistance (R_G), algebraic connectivity (μ_{N-1}) and ratio of μ_1/μ_{N-1} , have been applied to functional brain networks [27]. Is it redundant to consider these widely studied metrics to relate network property/metric to network functioning? Which set of metrics at least have to be considered? The understanding of the relation between network properties/metrics enables neuroscientists to discover the most relevant topological features/metrics that may characterize a certain brain disease or function [87].

This chapter investigates the correlations between structural and spectral network metrics, aiming to identify a small representative set of metrics by including only one metric from each sub-set of mutually strongly dependent metrics. Metric correlation was studied via the Pearson correlation coefficient¹ between network metrics in real-world networks in [88]. However, that approach did not address the following challenges. First, the correlation between network metrics is topology dependent. Consider, for example, the correlation between the average shortest path length $E[H]$ and the number of nodes N . The average shortest path length $E[H]$ is independent of the size N in the class of dense Erdős-Rényi random graphs but is positively correlated with N in D-dimensional lattices by $E[H] \sim (N^{1/D} \cdot \frac{D}{3})$. Thus, the correlation between two metrics can be different in different types of networks. In other words, the representative set of metrics can be different for different types of networks. Second, most network metrics are correlated with the number of nodes N (or even the number of links L) of a network. This introduces the difficulties in comparing networks [51, 89]. When the set of networks, like the set of real-world networks studied in [88] are of different size N , two metrics may seem to be strongly correlated, simply because they are both correlated with the size N . We approach these difficulties in three steps. (i) We consider three network models: the Erdős-Rényi random graphs [1] with a binomial degree distribution, the Barabasi-Albert

¹Pearson correlation coefficient is a measure of the extent of linear correlation between two variables.

graphs [7] with a power-law degree distribution and the Watts-Strogatz small-world graphs [6] where most nodes have the same degree. This allow us to understand the influence of the degree distribution on metric correlations. Although other network properties may as well influence metric correlations, we start with the degree distribution since it is the most studied and usually the easiest to obtain in most complex networks. (ii) We consider metric correlations in the instances of each model with a given size N and a given link density p . Initial results have been discussed in [42]. Here, we further explore how the metric correlations change with network parameters N and p to obtain the metric correlation pattern in each network model. (iii) The metric correlations are explored in the functional brain network of healthy subjects, which have the same network size N .

Sections 2.1.1 and 2.1.2 have introduced the network metrics [90] that we will explore in this chapter. The Pearson correlation coefficients between network metrics are computed in a large number of network instances of each model with various parameters in Section 3.3. The metric correlation patterns in network models as well as the corresponding representative sets of metrics are considered important contributions of this chapter. Surprisingly, the large set of metrics that we considered can be sufficiently represented by a small number of metrics in Erdős-Rényi random graphs, Barabasi-Albert graphs and Watts-Strogatz small-world graphs. The analytic relations between network metrics, presented in Section 3.2, support the correlations discovered via numerical experiments in Section 3.3.

Finally, we study metric correlations in the functional brain networks of healthy subjects in Section 3.4. We discuss how to derive unweighted functional brain networks via fixed threshold or via fixed average degree [89]. First, with the fixed threshold, the unweighted functional brain networks are shown to follow approximately binomial degree distribution. Interestingly, the metric correlation pattern in the studied functional brain networks is consistent with what we found in the Erdős-Rényi random graph model. Second, with the fixed average degree or link density, the degree distribution of unweighted functional brain networks has a heavy tail. The metric correlation patterns of this two types of unweighted functional brain networks are different whilst their degree distributions differ.

Our results suggest that a) the representative set of network metrics can indeed be smaller than the originally considered set; b) the average distance and the clustering coefficient, the most studied metrics so far especially in neuroscience, are strongly correlated, thus, redundant; c) spectral metrics² are only studied recently in the context of complex networks. However, at least one spectral metric appears in the representative set, suggesting the importance of spectral metrics in network characterizations. When we study a class of graphs with a given degree distribution, these networks can be possibly characterized by a small representative set of metrics instead of by the originally considered set. However, in the study of a specific complex network, the representative set at least has to be considered so as not to neglect any important network properties. In empirical networks studies, we probably don't know the true underlying topology but partial network properties. Our understanding of the dependency of metric correlation pattern on network properties opens up the possibility to use metric correlation as a topology diagnostic for real networks.

²Spectral metrics are those involving in the eigenvalue computations, such as spectral radius (λ_1), effective graph resistance (R_G), algebraic connectivity (μ_{N-1}) and ratio of μ_1/μ_{N-1} .

3.2 Analytic Relations Between Network Metrics

In this section, we will analytically derive relations between the network metrics introduced in Sections 2.1.1 and 2.1.2. Relations that have been proved in the literature will be presented as well. These analytic relations partially explain the observations of numerical experiments in Section 3.3.

3.2.1 General Relations

Lemma 3.2.1. *In any connected graph, for $N > 1$, the effective graph resistance R_G obeys*

$$\frac{R_G}{(N-1)^2} \geq \frac{1}{E[D]} \geq \frac{1}{\lambda_1} \quad (3.1)$$

Proof. From [14, p. 68], the sum of Laplacian eigenvalues equals

$$\sum_{j=1}^{N-1} \mu_k = 2L,$$

so that, for any graph with $L > 0$,

$$\frac{1}{\sum_{j=1}^{N-1} \mu_k} = \frac{1}{2L}. \quad (3.2)$$

The Jensen's inequality states that, if $f(x)$ is a convex function, (see [16, sec 5.2])

$$f(E[X]) \leq E[f(X)]. \quad (3.3)$$

Since $f(x) = 1/x$ is convex when $x > 0$, and nice in a connected graph, $\mu_k > 0$, for $1 \leq k \leq N-1$, application of (3.3) to the left-hand side of (3.2) yields

$$\frac{1}{\frac{1}{N-1} \sum_{j=1}^{N-1} \mu_k} \leq \frac{1}{N-1} \sum_{j=1}^{N-1} \frac{1}{\mu_k},$$

from which we obtain

$$\sum_{j=1}^{N-1} \frac{1}{\mu_k} \geq \frac{(N-1)^2}{2L}, \quad (3.4)$$

invoking the definition (2.6) of the effective graph resistance,

$$\frac{R_G}{(N-1)^2} \geq \frac{N}{2L} = \frac{1}{E[D]}. \quad (3.5)$$

We note that (3.5) is another derivation of the inequality (7.25) in [14]. Using the classical bound [14] of the spectral radius $\lambda_1 \geq E[D]$, we arrive by combining $\frac{1}{\lambda_1} \leq \frac{1}{E[D]}$ and (3.5) at (3.1). \square

Inequality (3.1) supports the negative correlation between the effective graph resistance and the spectral radius, if the inequality is close to an equality.

Relation 1 For any connected graph,

$$\lambda_1 \geq \frac{N_3}{N_2} \quad (3.6)$$

Proof. see [40]. □

Restrepo *et al.* [38] have shown that the spectral radius λ_1 can be approximated by N_3/N_2 . Hence, the inequality (3.6), proves the strong correlation $\rho(\lambda_1, \frac{N_3}{N_2}) \approx 1$. The equality occurs in regular graphs, so the more irregular a graph is, the worse is (3.6)

Relation 2 In any connected graph,

$$\frac{\mu_1}{\mu_{N-1}} = \frac{1}{\mu_{N-1}} \cdot \mu_1 > \frac{1}{\mu_{N-1}} \quad (3.7)$$

Since $\mu_1 \geq \frac{N}{N-1} D_{\max} > 1$ (see [14]).

The inequality (3.7) supports the strong negative correlation between $\frac{\mu_1}{\mu_{N-1}}$ and μ_{N-1} .

Relation 3 It is immediate from the Jensen's inequality (3.3) that $\frac{1}{E[H]} \leq E[\frac{1}{H}]$.

Hence, the average hopcount $E[H]$ is negatively correlated with the global efficiency $E[\frac{1}{H}]$.

Relation 4 As proved in [14, p. 207],

$$R_G \leq \binom{N}{2} E[H] \quad (3.8)$$

The inequality (3.8) supports the positive correlation between the effective graph resistance and the average hopcount.

Relation 5 In any Erdős-Rényi random graph or approximately in Bárábasi-Albert graphs

$$\lambda_1 \geq \kappa \quad (3.9)$$

Proof. The assortativity of an Erdős-Rényi random graph equals to 0, as proved in [17, 40]. Via (2.3), we have

$$N_2^2 = N_1 N_3,$$

where, $N_k = u^T A^k u$ are the total numbers of walks with k hops. Using

$$E[D] = \frac{N_1}{N}$$

and

$$E[D^2] = \frac{N_2}{N},$$

we arrive to

$$\kappa = \frac{E[D^2]}{E[D]} = \frac{N_2}{N_1} = \frac{N_3}{N_2}.$$

The bound in (3.6) yields (3.9). \square

The spectral radius λ_1 can be approximated by its lower bound $\frac{N_3}{N_2}$, equivalently by κ in an Erdős-Rényi random graph (if p is high) or B arabasi-Albert graph, which proves the strong correlation between λ_1 and κ .

3.2.2 Analytic Relations in Erdős-R enyi Random Graphs

Relation 6 In Erdős-R enyi random graphs $G_p(N)$, we have

$$E[H] \approx \begin{cases} 2 - C_G, & \text{when } p = 1 - o(\epsilon) \text{ or } N \text{ is sufficiently large} \\ \frac{\log N}{\log N + \log C_G}, & \text{when } Np = \varsigma \text{ (}\varsigma \text{ is a constant and } N \text{ is sufficiently large)} \end{cases}$$

Proof. The average clustering coefficient of an Erdős-R enyi random graph $G_p(N)$ is

$$E[C_G] = p. \quad (3.10)$$

When the link density p is large in Erdős-R enyi random graphs, the average hopcount [16] is

$$E[H] \simeq 2 - p. \quad (3.11)$$

From (3.10) and (3.11), we obtain

$$E[H] \simeq 2 - C_G.$$

In sparse Erdős-R enyi random networks, where $Np = \varsigma$ (ς is a constant), we have [16]

$$E[H] \approx \frac{\log N}{\log N + \log p} \approx \frac{\log N}{\log N + \log C_G}. \quad (3.12)$$

\square

Hence, Relation 6 explains why the average hopcount $E[H]$ is negatively and strongly correlated with the clustering coefficient C_G in Erdős-R enyi random graphs.

Relation 7 In Erdős-R enyi random graphs $G_p(N)$, we have

$$\begin{cases} E[\lambda_1] \approx (2 - E[H])(N - 2), & \text{when } p = 1 - o(\epsilon) \text{ or } N \text{ is sufficiently large} \\ E[H] \approx \frac{\log N}{\log E[\lambda_1]}, & \text{when } Np = \varsigma \text{ (}\varsigma \text{ is a constant and } N \text{ is sufficiently large)} \end{cases}$$

Proof. The average of spectral radius λ_1 in Erdős-R enyi random graphs can be expressed as [14]

$$E[\lambda_1] = p(N - 2) + 1 + O\left(\frac{1}{\sqrt{N}}\right). \quad (3.13)$$

When p is large, via (3.11), we have

$$E[\lambda_1] \approx (2 - E[H])(N - 2) + 1 = C_G(N - 2) + 1. \quad (3.14)$$

In sparse Erdős-Rényi random networks, using (3.12)

$$E[H] \approx \frac{\log N}{\log N + \log \frac{E[\lambda_1]}{N-2}} = \frac{\log N}{\log N - \log(N - 2) + \log E[\lambda_1]} \approx \frac{\log N}{\log E[\lambda_1]}.$$

□

Relation 7 supports the strong negative correlation between the spectral radius and the average hopcount. In dense Erdős-Rényi random graphs, (3.14) shows the strong positive correlation between the spectral radius and the clustering coefficient.

3.2.3 Analytic Relations in Barabasi-Albert Graphs

Lemma 3.2.2. *For Barabasi-Albert graphs, it holds, that,*

$$\lambda_1 \geq 2C_G^{-\frac{1}{3}} \quad (3.15)$$

Proof. In the Barabasi-Albert graph, when $m = 4$, the clustering coefficient C_G decreases [91] with the network size N as

$$C_G \approx N^{-\frac{3}{4}}. \quad (3.16)$$

In addition, when the number of added links m is equal to the starting number of vertices m_0 , the maximum degree of the Barabasi-Albert can be given as

$$D_{\max} = mN^{\frac{1}{2}}. \quad (3.17)$$

The lower bound of the spectral radius λ_1 can be given as the square root of the network's largest degree D_{\max} (see [14, art. 54, p. 55]), as

$$\lambda_1 \geq \sqrt{D_{\max}} = \sqrt{m}N^{\frac{1}{4}}. \quad (3.18)$$

Hence, we arrive at (3.15). □

Inequality (3.15) supports that the large Pearson correlation coefficient $\rho(\lambda_1, C_G)$ in Figure A.3.

Relation 8 In Barabasi-Albert graphs,

$$E[H] < \frac{1}{C_G} \quad (3.19)$$

Proof. In Barabasi-Albert graphs, the average hopcount approximates [91]

$$E[H] \approx \frac{\ln N}{\ln \ln N}. \quad (3.20)$$

When $N > 4$,

$$\frac{\ln N}{\ln \ln N} < N^{\frac{3}{4}}.$$

with (3.16), we obtain (3.19). □

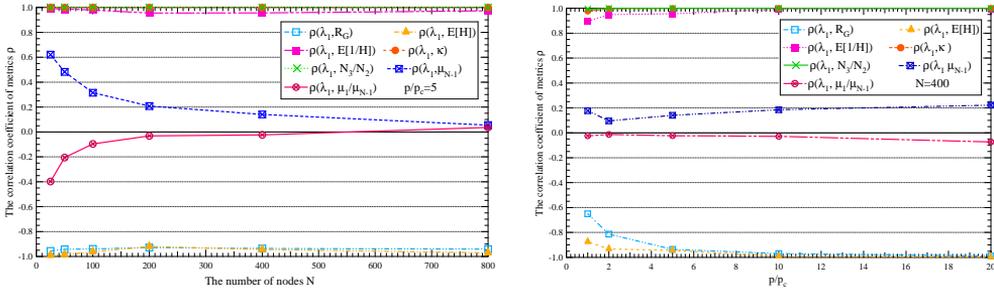


Figure 3.1: The correlation between the spectral radius and other metrics of Erdős-Rényi random graphs

The inequality (3.19) supports the simulation results in the left bottom diagram of Figure 3.1 and right bottom diagram of Figure A.3 that the average hopcount $E[H]$ and the clustering coefficient C_G are negatively correlated.

3.3 Metric Correlations in Network Models

In this section, we compute the Pearson correlation coefficient $\rho(i, j)$ between any two metrics i and j defined in Sections 2.1.1 and 2.1.2 in a large number of network instances of Erdős-Rényi random graphs, Barabasi-Albert graphs as well as Watts and Strogatz small-world graphs, which are introduced in Section 2.2. The matrix ρ is called the correlation matrix. The absolute value $0 \leq |\rho(i, j)| \leq 1$ characterizes the strength of the correlation between the corresponding metrics i and j . If $|\rho(i, j)|$ is close to zero, the two metrics are almost uncorrelated whereas a $|\rho(i, j)|$ close to 1 implies a strong correlation. We don't explore further the sign of $\rho(i, j)$ which reflects whether the correlation is positive or negative, because it is the strength $|\rho(i, j)|$ indicates to which extent a metric can be predicted from the other. Furthermore, we investigate how metric correlations change with network parameters, specifically, the network size N and the link density p of Erdős-Rényi random graphs and the network size N of Barabasi-Albert graphs and Watts and Strogatz small-world graphs.

3.3.1 Erdős-Rényi Random Graphs

The correlation $\rho(i, j)$ between any two metrics is computed in the 10^3 realizations of the Erdős-Rényi random graph $G_p(N)$, where $p = 5p_c$ and $N = 25, 50, 100, 200, 400, 800$. This allows us to explore how the metric correlation $\rho(i, j)$ evolves with the network size N . Similarly, the metric correlations are also computed in the 10^3 instances of the Erdős-Rényi random graph $G_p(N)$ where $N = 400$ and $p = \alpha p_c$ with $\alpha \in [1, 2, 5, 10, 20]$ to examine the influence of link density p on metric correlations.

Figure 3.1 illustrates the Pearson correlation coefficient between λ_1 and other metrics for different size N and link density p . The correlation between λ_1 and κ , N_3/N_2 , $E[1/H]$ are positive and strong over all network sizes and link densities: $\rho(\lambda_1, \kappa) \simeq \rho(\lambda_1, N_3/N_2) \simeq \rho(\lambda_1, E[1/H]) \simeq 1$. These three positive correlations are supported by analytic relations

Relation 5, Relation 1, Relation 7 and 3 respectively. The correlation between λ_1 and R_G , $E[H]$ are negative and strong $\rho(\lambda_1, R_G) \simeq \rho(\lambda_1, E[H]) \simeq -1$ with the only condition that the link density should not be too small, thus $p \geq 4p_c$. These observations can be analytically explained by Lemma 3.2.1 and Relation 7, respectively. On the other hand, λ_1 tends to be independent or weakly correlated with μ_{N-1} and μ_1/μ_{N-1} : $|\rho(\lambda_1, \frac{\mu_1}{\mu_{N-1}})| < 0.2$ and $|\rho(\lambda_1, \mu_{N-1})| < 0.2$, when the network size is large. Similarly, the correlation coefficient between any other two metrics as a function of the network size as well as the link density has been given in Appendix A.

We consider two metrics i and j as uncorrelated if $|\rho(i, j)| \leq 0.2$ and as strongly correlated if $|\rho(i, j)| \geq 0.7$, because most of the correlations $|\rho(i, j)|$ are in the range $[0, 0.2]$ and $[0.7, 1]$. Since the correlation $\rho(i, j)$ changes with the network size N and link density p , we claim two metrics are strongly correlated (or independent) if $|\rho(i, j)| \geq 0.7$ (or $|\rho(i, j)| \leq 0.2$) holds for a certain range of N and p and we will record the condition on N and p , under which this strong (or weak) correlation is observed. In this way, we could obtain the metric correlation pattern based on results in Figures 3.1, A.1 and A.2 and the correlation pattern is represented as a graph in Figure 3.2.

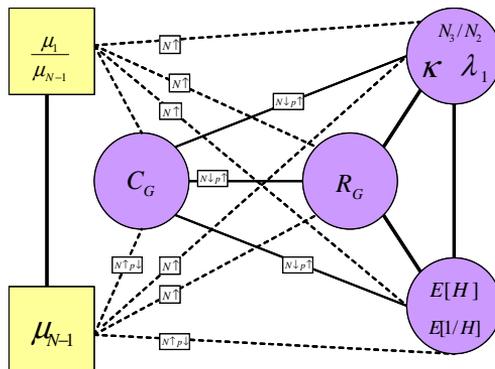


Figure 3.2: The metric correlation pattern in Erdős-Rényi random graphs.

In correlation patterns, each node represents a metric. Two nodes are connected by a solid (or dotted) line if the corresponding metrics are strongly correlated (or independent) with $|\rho(i, j)| \geq 0.7$ (or $|\rho(i, j)| \leq 0.2$). The boxes along the links specify the conditions, if there are, under which the strong or weak correlation has been observed. The up (down) arrow represents the value of N or p is large (or small). For example, when the size N of the network is small and the link density p is large, the correlation between C_G and R_G is strong. Metric λ_1 , κ and N_3/N_2 are strongly mutually correlated and they have the same correlation coefficient with any other metric. Hence, we condense these three metrics into one node. The same holds for $E[H]$ and $E[1/H]$.

Interestingly, our approach allows us to cluster the metrics into 2, as marked by different shapes of nodes, when the size N of network is small and link density p is large. Within each cluster, metrics are mutually strongly correlated (nodes are fully connected by solid lines). Moreover, any two metrics from different clusters are independent (any two nodes with different shapes are linked by a dotted line). Thus, it is sufficient to characterize an Erdős-Rényi random graph by two metrics, each from a different cluster, instead of by the 9 metrics studied. In large and sparse networks, the correlation between

clustering coefficient and the metrics in circle is not strong any more. In that case, the representative set should contain three metrics: the clustering coefficient C_G , one metric in rectangle ($\frac{\mu_1}{\mu_{N-1}}$ or μ_{N-1}) and one metric from the circle cluster ($R_G, \lambda_1, \kappa, N_3/N_2, E[H]$ or $E[1/H]$).

Those strong and weak metric correlations observed in Erdős-Rényi random graphs are supported by the analytic relations between metrics in Section 3.2 as shown in Table 3.1. Note that, in tables, the Lemma and Relation in this section are written as L and R for short, *e.g.* Lemma 3.2.1 is written as L1, and Relation 1 is represented by R1.

3.3.2 Barabasi-Albert Power-law Graphs

The Pearson correlation coefficients between any two metrics are computed in 10^3 BA graphs with each given set of parameters m and N , where $m = 4$ and $N = 25, 50, 100, 200, 400, 800$. The metric correlations as a function of the network size N are illustrated in Figure A.3 in Appendix A. Based on these results, we obtain the metric correlation pattern in Barabasi-Albert graphs as depicted in Figure 3.3.

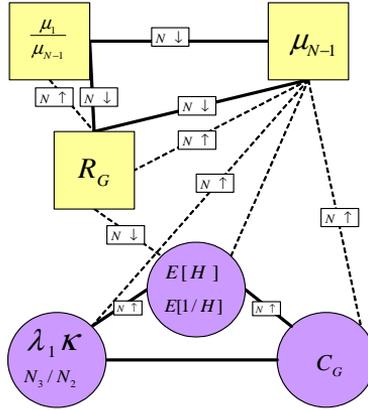


Figure 3.3: Metric correlation pattern in Barabasi-Albert graphs.

We first consider the case when the network size N is large. The metrics in circle are mutually strongly correlated with each other and they are uncorrelated with the other metrics. Hence, one and only one metric in the circle cluster should be included in the representative set of metrics. On the other hand, the three metrics in rectangle, $\frac{\mu_1}{\mu_{N-1}}$, μ_{N-1} and R_G are not strongly correlated in large networks. Thus, these three metrics should all be included in the representative set. In summary, four metrics, the three marked rectangle and one in circle, suffice to represent a power-law graph with a large N . When the network size is small, it is sufficient to characterize a network by 3 metrics κ (λ_1 or N_3/N_2), $E[H]$ (or $E[1/H]$) and one metric from the rectangle cluster ($\frac{\mu_1}{\mu_{N-1}}$, μ_{N-1} and R_G). This observed correlation pattern is supported by the analytic relations between metrics in Section 3.2 as summarized in Table 3.2.

Table 3.1: Verification of the observed metric correlations in Erdős-Rényi random graph by analytic relations between metrics in Section 3.2.

$\rho(i, j)$	R_G	μ_{N-1}	λ_1	$\frac{\mu_1}{\mu_{N-1}}$	$E[H]$	$E[\frac{1}{H}]$	C_G	κ	N_3/N_2
R_G	1		L1		R4	R3, R4	R4, R6	L1, R5	L1, R1
μ_{N-1}		1		R2					
λ_1	L1		1		R7	R3, R7	R6, R7	R5	R1
$\frac{\mu_1}{\mu_{N-1}}$				1					
$E[H]$	R4		R7		1	R3	R6	R5, R7	R1, R7
$E[\frac{1}{H}]$	R3, R4		R3, R7		R3	1	R3, R6	R3, R5, R7	R1, R3, R7
C_G	R4, R6		R6, R7		R6	R3, R6	1	R5, R6, R7	R1, R6, R7
κ	L1, R5		R5		R5, R7	R3, R5, R7	R5, R6, R7	1	R1
N_3/N_2	L1, R1		R1		R1, R7	R1, R3, R7	R1, R6, R7	R1	1

Table 3.2: Verification of the observed metric relations in Barabasi-Albert graph by analytic relations between metrics in Section 3.2.

$\rho(i, j)$	R_G	μ_{N-1}	λ_1	$\frac{\mu_1}{\mu_{N-1}}$	$E[H]$	$E[\frac{1}{H}]$	C_G	κ	N_3/N_2
R_G	1		L1		R4	R3, R4	L1 and L2	L1, R5	L1, R1
μ_{N-1}		1		R2					
λ_1	L1		1		L1 and R4	L1, R3, R4	L2	R5	R1
$\frac{\mu_1}{\mu_{N-1}}$				1					
$E[H]$	R4		L1, R4		1	R3	R8	L1, L4, L5	L1, R1, R4
$E[\frac{1}{H}]$	R3, R4		L1, R3, R4		R3	1	R3, R8	L1, R3, R4, R5	L1, R1, R3, R4
C_G	L1, R2		L2		R8	R3 and R8	1	L2, R5	L2, R1
κ	L1, R5		R5		L1, R4, R5	L1, R3, R4, R5	L2, R5	1	R1
N_3/N_2	L1, R1		R1		L1, R1, R4	L1, R1, R3, R4	L2, R1	R1	1

3.3.3 Watts-Strogatz Small-world Graphs

We calculate the Pearson correlation coefficients between any two metrics in 10^3 instances of the WS model with each given set of parameters k , p_r and N , where $k = 6$, $p_r = 0.01$ and $N = 200, 400, 800, 1000, 2000$. The correlation coefficient between any other two metrics as a function of the network size has been given in Figure A.4 in Appendix A. With these results, we obtain the metric correlation pattern in Watts-Strogatz small-world graphs in Figure 3.4.

We first consider the case when the network size N is small. In the circle cluster, metrics, *i.e.* R_G , $\kappa (N_3/N_2)$, $E[H]$ ($E[1/H]$) C_G and μ_{N-1} , are mutually strongly correlated with each other, except that $\frac{\mu_1}{\mu_{N-1}}$ is uncorrelated with $\kappa (N_3/N_2)$ and C_G . Furthermore, λ_1 is only correlated with $\kappa (N_3/N_2)$, and the assortativity ρ_D in triangle is not correlated all other metrics in networks. Hence, four metrics should be included in the representative set: $\rho_D, \lambda_1, \frac{\mu_1}{\mu_{N-1}}$ and another metric in the circle cluster. When the network size is large, it is still enough to characterize a network by 4 metrics $\lambda_1, \frac{\mu_1}{\mu_{N-1}}$ (or μ_{N-1}), one from other metrics in circle and ρ_D .

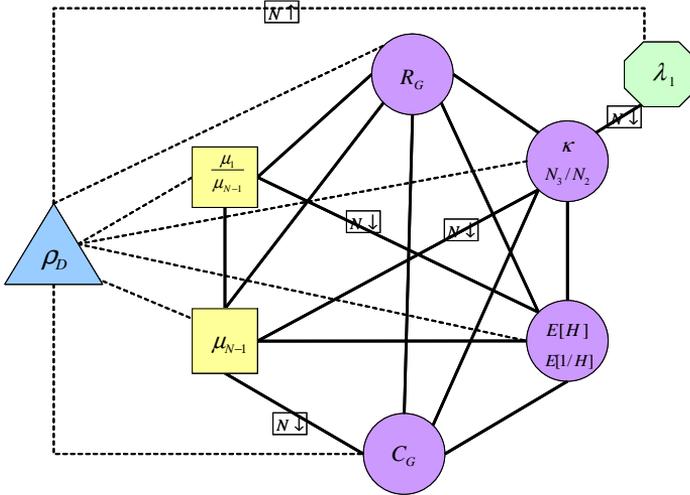


Figure 3.4: Metric correlation pattern in Watts-Strogatz small-world graphs.

The different metric correlation patterns observed in Erdős-Rényi random graphs (Figure 3.2), Bárábasi-Albert graphs (Figure 3.3) and Watts-Strogatz small-world graphs (Figure 3.4) reflect that metric correlations depend on the graph, via the adjacency matrix, e.g. the degree distribution. Both network models can be characterized by a small set of 3 or 4 metrics instead of by the 9 metrics studied. Surprisingly, the average hop-count $E[H]$ and clustering coefficient C_G , the most studied metrics so far especially in neuroscience turn out to be strongly correlated³. On the other hand, spectral metrics, which are investigated only recently, seem to be essential in network characterizations. At least, spectral metrics always appear in the representative set in these three network models.

³Note that this strong correlation is expected in a class of networks with the same degree distribution as in the three models, but not in graphs with different degree distributions.

3.4 Metric Correlations in Functional Brain Networks

It has become clear that properties (reflected by metrics) of brain networks may predict brain functioning such as cognitive performance [25, 26, 27, 28]. Moreover, brain networks show systematic changes during development under genetic control and reveal characteristic patterns of disruption in various types of neurological disease. In order to understand which network property/metric is the most relevant to a certain brain functioning, it is essential to understand the relations between network properties. In this section, we investigate the metric correlation pattern in the functional brain networks of 22 healthy subjects. The correlation pattern observed in functional brain networks will be compared with what we discovered in network models.

The concept of functional connectivity refers to the statistical dependencies between physiological time series recorded in various brain areas, and is thought to reflect communication between several brain areas [92]. Magnetoencephalography (MEG), a recording of the brains magnetic activity, is a method used to assess functional connectivity within the brain. A functional brain network is created by regarding each MEG channel as a node, and the functional connectivity between each pair of channels represents a link whose weight reflects the strength of the connectivity or correlation. Correlations between the time series of the channels were analyzed with the synchronization likelihood (SL), a non-linear measure of statistical interdependencies [93, 94]. The functional brain networks were measured in 22 healthy people, whose mean age was 63.6, by the medical ethics committee of the VU Medical Center, Amsterdam. More information about the data can be found in [95]. Brain activity was measured by MEG with $N = 149$ recording channels. With MEG, a loss of upper alpha-, beta-, and gamma-band synchronization could be demonstrated in patients with Alzheimer’s disease, both during an eye-closed state as well as during an eye-open state. In these studies, the 13– 30 Hz beta band showed the most consistent abnormalities in the subjects [66]. Hence, we focus on the 13– 30 Hz band. The functional brain network was measured four times for each person in 13– 30 Hz (beta) band. In total, we have 88 functional brain networks. Each network is a weighted complete graph, where each link weight $0 \leq w_{ij} \leq 1$.

Since we focus on metrics correlations in unweighted networks, each weighted network is transformed to an unweighted network mapped as the union of links whose link weight is above a threshold T . There are many methods to choose the threshold T [89]. In this section, we use either a fixed threshold T for all networks or a network dependent T which makes the link density p of all unweighted networks similar. The metric correlation patterns are studied in the two resulting classes of networks with different degree distributions.

3.4.1 Metric Correlations in Unweighted Functional Brain Networks Transformed with Fixed Threshold T

Stam *et al.* [66] investigated the clustering coefficients and the shortest path lengths as a function of the threshold T on functional brain networks, and gave some suggestions on how to set the threshold T . When the threshold T is small, the corresponding unweighted functional brain networks are almost fully connected. If T is large, the unweighted network tends to be disconnected. Regarding to the studied functional brain networks, we find that when $T \in [0.001, 0.019]$, the corresponding unweighted functional brain networks are

connected. When $T \geq 0.02$, not all the networks are connected. In summary, to avoid disconnected and fully connected graphs, we choose $T = 0.019$ to transform each weighted functional brain network into an unweighted one, where we calculate the correlation between those metrics mentioned above. As shown in Figure 3.5, the degree distribution of the unweighted functional brain networks is close to a binomial distribution, as in Erdős-Rényi random graphs.

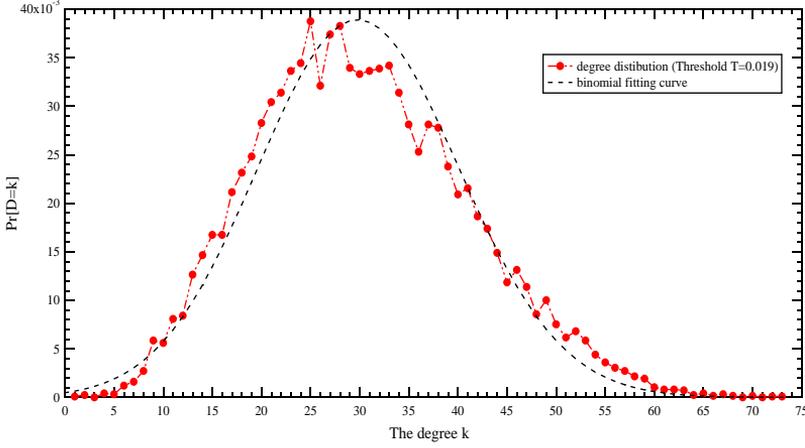


Figure 3.5: The degree distribution of unweighted functional brain networks (fixed $T = 0.019$) as well as its binomial curve fitting.

The Pearson correlation coefficients between any two metrics in the unweighted functional brain networks are shown in Table 3.3.

Table 3.3: Correlation coefficients between metrics in unweighted functional brain networks (healthy people, beta-band, fixed threshold $T = 0.019$)

$\rho(i, j)$	R_G	μ_{N-1}	λ_1	$\frac{\mu_1}{\mu_{N-1}}$	$E[H]$	$E[\frac{1}{H}]$	C_G	ρ_D	κ	N_3/N_2
R_G	1.00	-0.77	-0.94	0.31	0.98	-0.97	-0.91	-0.33	-0.94	-0.94
μ_{N-1}	-0.77	1.00	0.66	-0.73	-0.73	0.72	0.60	0.23	0.67	0.67
λ_1	-0.94	0.66	1.00	-0.11	-0.97	0.98	0.92	0.37	1.00	1.00
$\frac{\mu_1}{\mu_{N-1}}$	0.31	-0.73	-0.11	1.00	0.20	-0.19	-0.16	-0.16	-0.12	-0.11
$E[H]$	0.98	-0.73	-0.97	0.20	1.00	-1.00	-0.88	-0.26	-0.97	-0.97
$E[\frac{1}{H}]$	-0.97	0.72	0.98	-0.19	-1.00	1.00	0.90	0.27	0.99	0.99
C_G	-0.91	0.60	0.92	-0.16	-0.88	0.90	1.00	0.46	0.91	0.92
ρ_D	-0.33	0.23	0.37	-0.16	-0.26	0.27	0.46	1.00	0.31	0.34
κ	-0.94	0.67	1.00	-0.11	-0.97	0.99	0.91	0.31	1.00	1.00
N_3/N_2	-0.94	0.67	1.00	-0.11	-0.97	0.99	0.92	0.34	1.00	1.00

The corresponding correlation pattern is depicted in Figure 3.6. Since all the functional brain networks have the same size N , no condition is associated to any correlation or dependency. Instead, we place the correlation coefficient along each pair of metrics.

Both the functional brain networks and Erdős-Rényi random graphs possess a binomial

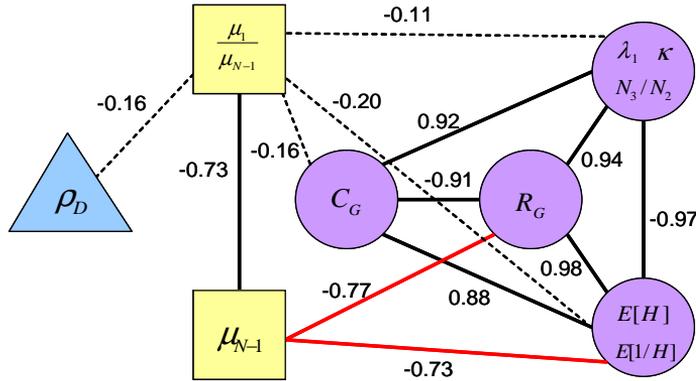


Figure 3.6: The metric correlation pattern of unweighted functional brain networks (fixed threshold T).

degree distribution. Interestingly, the metric correlation patterns observed in both types of graphs as given in Figures 3.2 and 3.6 are consistent with each other: a) the same set of metrics in the circle cluster as well as in the rectangle cluster; b) the size of functional brain networks $N = 149$ is small. In small Erdős-Rényi random graphs, the clustering coefficient is strongly correlated with the other metric in circle, moreover, $\rho(\mu_{N-1}, R_G)$ and $\rho(\mu_{N-1}, E[H])$ are not small, the same as observed in the functional brain networks. The only difference is that the assortativity ρ_D is usually considered in real-world complex networks but not in network models where $\rho_D \rightarrow 0$. Actually, functional brain networks are assortative $\rho_D > 0$. The representative set ρ_D , $\frac{\mu_1}{\mu_{N-1}}$ and R_G is preferable due to the independence between $\frac{\mu_1}{\mu_{N-1}}$ and metrics in circle as well as the strong correlation between R_G and many other metrics. The consistency in metric correlation patterns (Figures 3.2 and 3.6) in networks with a same degree distribution verifies the crucial influence of network property, specially the degree distribution, on the relations between metrics. Importantly, both the degree distribution and the metric correlation pattern suggest this particular class of MEG unweighted functional brain networks with a fixed threshold are almost Erdős-Rényi graphs like, and not small-world.

3.4.2 Metric Correlations in Unweighted Functional Brain Networks Transformed with Fixed Link Density p

Van Wijk *et al.* [89] claimed that graph measures can be influenced by the number N of nodes and the link density p . It is easier to compare networks, such as the functional brain networks, with the same size N and link density p . Thus, we choose a threshold T for each weighted network such that when the average degree of the corresponding unweighted network equals to a given constant value. In this section, we choose the threshold such that the average degree equals to 15. Each unweighted network follows almost the same degree distribution. The degree distribution of the unweighted functional brain networks is shown in Figure 3.7, which has a heavy tail. It differs from the degree distribution of any of the three network models mentioned in Section 3.3. The Pearson correlation coefficients between any two metrics in the unweighted functional brain networks are shown in Table 3.4. The corresponding correlation pattern is depicted in Figure 3.8.

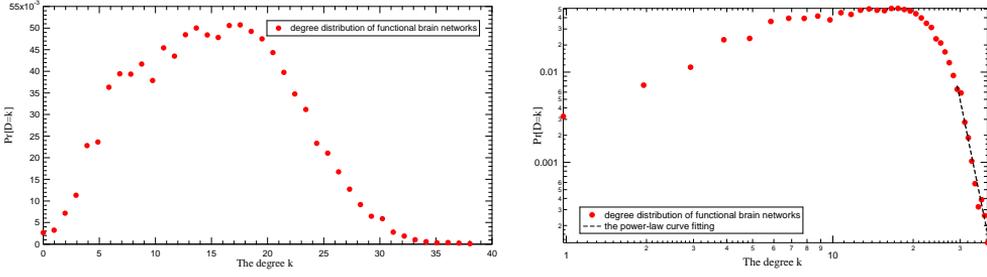


Figure 3.7: Degree distribution of unweighted functional brain networks (fixed link density p).

Table 3.4: Correlation coefficients between metrics in unweighted functional brain networks (healthy people, beta-band, same size (L and N))

C_{ij}	R_G	μ_{N-1}	λ_1	$\frac{\mu_1}{\mu_{N-1}}$	$E[H]$	$E[\frac{1}{H}]$	C_G	ρ_D	κ	N_3/N_2
R_G	1.00	-0.83	0.61	0.89	0.73	-0.69	0.54	0.47	0.64	0.68
μ_{N-1}	-0.83	1.00	-0.51	-0.85	-0.73	0.72	-0.63	-0.51	-0.47	-0.52
λ_1	0.61	-0.51	1.00	0.56	0.48	-0.47	0.44	0.55	0.85	0.94
$\frac{\mu_1}{\mu_{N-1}}$	0.89	-0.85	0.56	1.00	0.67	-0.61	0.46	0.41	0.55	0.59
$E[H]$	0.73	-0.73	0.48	0.67	1.00	-0.99	0.90	0.76	0.30	0.42
$E[\frac{1}{H}]$	-0.69	0.72	-0.47	-0.61	-0.99	1.00	-0.94	-0.81	-0.25	-0.39
C_G	0.54	-0.63	0.44	0.46	0.90	-0.94	1.00	0.85	0.19	0.34
ρ_D	0.47	-0.51	0.55	0.41	0.76	-0.81	0.85	1.00	0.19	0.39
κ	0.64	-0.47	0.85	0.55	0.30	-0.25	0.19	0.19	1.00	0.97
N_3/N_2	0.68	-0.52	0.94	0.59	0.42	-0.39	0.34	0.39	0.97	1.00

The metrics can be divided into three groups, which are shaped rectangle, circle and triangle. Metrics in the same shape are strongly correlated with each other, while, metrics in different shapes are not strongly correlated. We could choose one metric from each shape group, therefore, three representative metrics can characterize the unweighted functional brain networks.

3.5 Chapter Conclusions

In this chapter we have studied the correlations between widely studied structural and spectral metrics in functional brain networks as well as in three classical complex network models, namely Erdős-Rényi, Bárábasi-Albert graphs and Watts-Strogatz small-world graphs. The metric correlation pattern in each of the three classes of graphs illustrates the strong correlations and independencies between metrics, which indicates the possibility to determine a small set of representative metrics by including only one metric from each sub-set of mutually strongly dependent metrics. This representative set of metrics tends to characterize both sufficiently and effectively a class of networks with a given degree distribution. When we study a specific network, however, the representative set at least has to be considered so that important network properties will not be overlooked. Most of the metric correlations observed so far are supported/explained analytically by

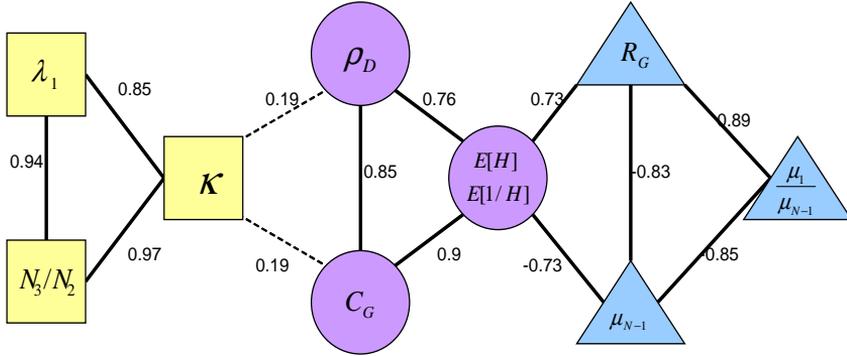


Figure 3.8: *The metric correlation pattern of unweighted functional brain networks (fixed link density p).*

theorems developed in this chapter as well as in the literature. Furthermore, graphs with a similar degree distribution such as the unweighted functional brain networks with fixed threshold T and Erdős-Rényi random graphs, tend to possess a similar metric correlation pattern, which verifies the influence of the degree distribution on metric relations. When the degree distribution of networks are different, the metric patterns are also different. Hence, a specific representative set of metrics should be determined for a particular degree distribution. Luckily, degree distribution is the simplest metric to compute. Finally, two observations are considered important especially for the applications of network science to other disciplines: i) the average distance and the clustering coefficient, the most studied metrics so far in neuroscience, are strongly correlated, thus, redundant; ii) spectral metrics, though only studied recently in the context of complex networks seem to be essential in network characterizations. Note that the metric correlations are studied and observed in networks with a similar degree distribution. These results can not simply be applied to a class of networks with different degree distributions.

CHAPTER 4

Correlations Between Centrality Metrics

“Measure what can be measured, and make measurable what cannot be measured.”

Galileo Galilei, 1564 - 1642

THE metrics studied in Chapter 3 are used to characterize the topology and the functions of a network. In this chapter, we aim to understand the correlations between centrality metrics, which have been proposed to rank the importance of nodes, so that we could approximate a high-complexity centrality metric by a strongly correlated low-complexity one. We first investigate the relation between the degree vector and the principal eigenvector. The average $E[x_1]$ is shown to decrease with the assortativity ρ_D . The difference between the principal eigenvector and the scaled degree eigenvector is smallest, when $\lambda_1 = \frac{N_2}{N_1}$, where N_k is the total number walks in the graph with k hops. Moreover, the principal eigenvector and the degree vector are uncorrelated when ρ_D is low in the power-law networks, albeit they are strongly correlated in other networks. Next we study the correlation between a number of centrality metrics (introduced in Section 2.1.3) by their Pearson correlation coefficient and their similarity in ranking of nodes. In addition to considering the widely used centrality metrics, we introduce a new centrality measure, the degree mass. The m th-order degree mass of a node is the sum of the degree of the node and the degree of its neighbors not further than m hops away. We find that the betweenness, the closeness and the components of the principal eigenvector are strongly correlated with the degree, the 1st-order degree mass and the 2nd-order degree mass, respectively, in both network models and real-world networks. We then theoretically prove that the Pearson correlation coefficient between the principal eigenvector and the 2nd-order degree mass is larger than that between the principal eigenvector and a lower order degree mass. Finally, we investigate the effect of the inflexible contrarians selected based on different centrality metrics in helping one opinion to compete with another in

the inflexible contrarian opinion (ICO) model. Interestingly, we find that selecting the inflexible contrarians based on the leverage, the betweenness, or the degree is more effective in opinion-competition than using other centrality metrics in all types of networks. This observation is supported by our previous observations, *i.e.*, that there is a strong linear correlation between the degree and the betweenness, as well as a high centrality similarity between the leverage and the degree.

4.1 Introduction

Recent research has explored social dynamics [60, 69, 11] by using complex networks in which nodes represent people/agents and links the associations between them. Such centrality metrics as degree and betweenness have been studied in dynamic processes [96, 21, 97, 98], such as opinion competition, epidemic spreading, and rumor propagation on complex networks. These studies used centrality metrics to identify influential nodes [96, 21, 97], such as the source nodes from which a virus spreads and the nodes with high spreading capacity, as well as to select which nodes are to be immunized when a virus is prevalent [98]. Numerous centrality metrics have been proposed. Degree, betweenness, closeness, and principal eigenvector are the most popular centrality metrics [96, 99, 54, 100, 101, 102, 103]. Several new centrality metrics have been introduced in a number of different fields recently. Kitsak *et al.* [21] studied the SIS and SIR spreading models on four real-world networks and proposed that the k -shell index is a better indicator for the most efficient spreaders (nodes) than degree or betweenness. Reference [22] proposes a new centrality metric—*leverage*—for identifying neighborhood hubs (the most highly-connected nodes) in functional brain networks. Leverage centrality identifies nodes that are connected to more nodes than their nearest neighbors. In addition to considering these widely-used centrality metrics, we here propose a new centrality metric, *degree mass*. The m th-order degree mass of a node is defined as the sum of the weighted degree of its m -hop neighborhood. If the degree of a node and of its neighbors are all high, the node has a high degree mass.

Centrality metrics have been compared in various networks, such as sampled networks, biological networks, food webs, and vocabulary networks in literature [96, 104, 105, 106, 107]. Comin *et al.* [96] compared the centrality metrics characterizing the performances of nodes in such dynamic processes as virus spreading. Kim and Jeong [104] compared the reliability of rank orders using centrality metrics in sampling networks. The correlations between centrality metrics have been studied in biological networks [105, 106]. However correlations between centrality metrics are still not well understood. If correlations between centrality metrics were better understood, we might be able to rank the nodes in a network by using the centrality metrics with a low computational complexity instead of the ones with a high computational complexity.

Van Mieghem *et al.* [108] have proved that to minimize the largest eigenvalue by removing a set of links or nodes is an NP-hard problem and have shown that the best strategy so far is based on the components of the principal eigenvector x_1 , which underlines the importance of the principal eigenvalue in characterizing the influence of link/node removal on the spectral radius. We first investigate the relation between x_1 and the degree vector/sequence¹ d , the computationally simplest and mostly studied property of a

¹The degree vector/sequence is composed of the degree of each node, following the same ordering as

network. The relation between the principal eigenvector and the degree vector is systematically investigated in networks with various degree distributions and degree correlations (also called the assortativity, which is introduced in Section 2.1.1).

To investigate the correlation between any two centrality metrics introduced in Sec. 2.1.3, we compute their Pearson correlation coefficient and their similarity in ranking nodes in both network models and real-world networks. In this work (i) we consider the Erdős-Rényi (ER) networks [1] and the scale-free (SF) networks [7, 13]. Studying these two network models allows us to understand how the degree distribution influences correlations between the centrality metrics. (ii) We further explore correlations in 34 real-world networks with differing numbers of nodes and links. (iii) We theoretically compare the Pearson correlation coefficients between the principal eigenvector and the degree masses.

Recently there has been considerable interest in understanding how two competing opinions [109, 74, 75, 110, 111] evolve in a population. In this work we apply our centrality metrics to an inflexible contrarian opinion (ICO) model [76] in which only two opinions (denoted A and B) exist, with the goal of helping one opinion (opinion B) as it competes with with the other opinion (opinion A). At the initial time, opinions are randomly assigned to all nodes (with a fraction f of nodes holding opinion A and a fraction $1 - f$ of nodes holding opinion B). At each step, each agent simultaneously and in parallel adopts the opinion of the majority of its nearest neighbors and itself, and if there is a tie, the agent does not change its opinion. After the system reaches a steady state, a fraction p_o of agents with opinion A is placed among the inflexible contrarians permanently holding opinion B , which can affect the opinion of their nearest neighbors. It is known that the size of the giant component of agents with opinion A can be decreased or even destroyed by the inflexible contrarians [76]. Li *et al.* [76] have selected the inflexible contrarians in ER and SF networks either randomly or based on degree. Here we choose inflexible contrarians using all the centrality metrics we have considered in both modelled networks and real-world networks. We compare the efficiencies of these centrality metrics in reducing the size of the largest opinion A cluster and find that strongly correlated centrality metrics have approximately the same efficiency in both modelled networks and real-world networks. Thus a high-complexity centrality metric could be approximated by a strongly correlated low-complexity centrality metric.

4.2 Analyzing the Correlation Between Degree and Principal Eigenvector

In view of the importance of the principal eigenvector in characterizing the influence of link/node on the spectral radius, in this section, we explore how the average $E[x_1]$ as well the variance of x_1 changes with the assortativity ρ_D when the degree vector, which may follow the degree distribution of network models or of real-world networks, remains the same. Moreover, we explore the difference and the Pearson correlation coefficient between the principal eigenvector and the degree vector, the simplest and mostly studied network metric, which as well provides important insights on under which condition the degree vector/sequence well approximates the principal eigenvector [107].

the principal eigenvector.

4.2.1 Properties of the Principal Eigenvector

Two types of degree distributions have been so far widely studied: the binomial and power-law degree distribution. The binomial degree distribution is a characteristic of an Erdős-Rényi random graph $G_p(N)$, which has N nodes and any two nodes are connected independently with a probability p . Such a random construction leads to a zero assortativity as proved in [40]. However, the class of graphs $G(N, p)$ with the same binomial degree distribution $\text{Prob}[D = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k}$ as Erdős-Rényi random graphs $G_p(N)$ and obtained, for instance, by degree-preserving rewiring feature an assortativity that may vary within a wide range. The power-law degree distribution $\text{Pr}[D = k] = ck^{-\alpha}$, where $c = 1/\sum_{k=1}^{N-1} k^{-\alpha}$ has been widely observed in real-world networks. Similarly, graphs with a given power-law degree distribution, for example, generated by the Barabási-Albert power model [7] can be altered by the degree-preserving rewiring to obtain different assortativity.

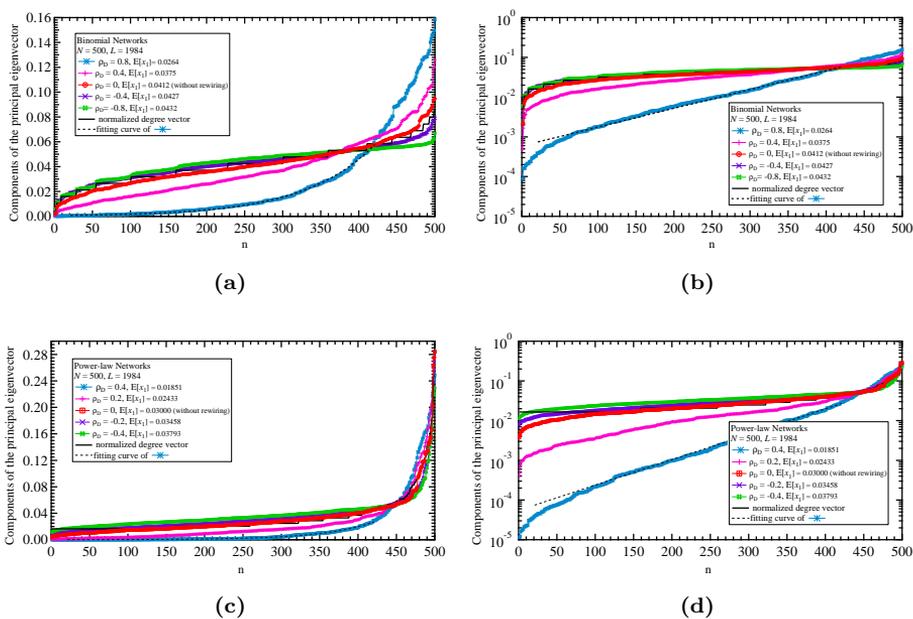


Figure 4.1: Components of the largest eigenvector in order. Images (a) (linear) (b) (semilogarithmic) are binomial graphs with different assortativity. Images (c) (linear) (d) (semilogarithmic) are power-law graphs with different assortativity.

We explore the principal eigenvector components (see Figure 4.1) as well as its average $E[x_1]$ (see Figure 4.2) in graphs with the same degree distribution (*i.e.* binomial or power-law) but with different assortativities ρ_D obtained by degree-preserving rewiring. Figure 4.1 shows that the variance of the principal eigenvector increases with assortativity ρ_D . Furthermore, as shown in Figure 4.2, $E[x_1]$ decreases with the increase of assortativity ρ_D . Similarly, we consider a set of 11 real-world networks. We apply degree-preserving rewiring to each real-world network to derive network instances with different assortativity. In other words, we derive a class of networks that possess the same degree distribution as a real-world network but different assortativities. Interestingly, we observe the same,

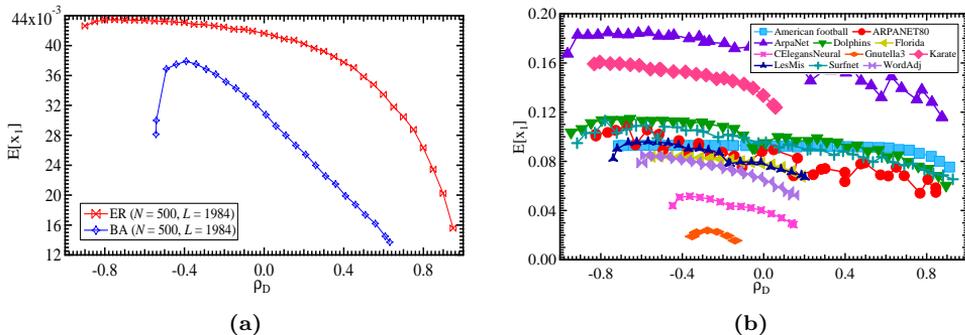


Figure 4.2: Average of the components of the principle eigenvector versus the assortativity. (a) in binomial and power-law graphs (b) in network instances derived from real-world networks via degree-preserving rewirings.

$E[x_1]$ decreases with increasing assortativity (see Figure 4.2(b)).

The decrease of $E[x_1]$ and the increase of the variance of the principal eigenvector components with increasing assortativity can be qualitatively explained as follows. As defined, the principal eigenvector x_1 corresponds to the largest eigenvalue λ_1 follows

$$\lambda_1(x_1)_j = \sum_{q=1}^N a_{jq}(x_1)_q, \quad (4.1)$$

where $a_{jq} = 1$ if q is a neighbor of node j , or else $a_{jq} = 0$. The j -th component of the principal eigenvector $(x_1)_j$ tends to be large if node j has a large degree (number of neighbors) or if the components corresponding to its neighbors are large. When ρ_D is large, high degree nodes prefer to link with other high degree nodes. In this case, a high degree node possesses a large number of neighbors, whose corresponding eigenvector components are again likely to be large, whereas a low degree node connects to a small number of neighbors, whose corresponding components tend to be small. Both a large variance in degree and a large assortativity ρ_D contribute to a large variance $Var[x_1]$ of the principal eigenvector x_1 . This explains why variance $Var[x_1]$ of x_1 increases with ρ_D and with a given assortativity, the power-law graphs have a larger $Var[x_1]$ than the binomial graphs (see Figure 4.1). Furthermore, since $Var[X] = E[X^2] - (E[X])^2$ and $x_1^T x_1 = 1$,

$$E[x_1] = \sqrt{\frac{1}{N} - Var[x_1]}, \quad (4.2)$$

Correspondingly, both a large variance in degree and a large assortativity ρ_D contribute to a small $E[x_1]$ of the principal eigenvector x_1 . Hence, $E[x_1]$ decreases with increasing ρ_D and tends to be smaller when the degree variance is larger. Moreover, considering Eq. (4.2), we can deduce the upper bound $E[x_1] \leq \frac{1}{\sqrt{N}}$.

Figure 4.1 compares as well the principal eigenvector x_1 with the normalized degree vector $\bar{d} = \frac{d}{\sqrt{d^T d}}$ in binomial graphs and power-law graphs ($N = 500, L = 1984$) with different assortativities. The components of x_1 and \bar{d} are plotted in the order of increasing

magnitude. The difference between x_1 and \bar{d} is affected by ρ_D , which will be further explored in the following part.

4.2.2 Relation Between the Degree Vector and the Principal Eigenvector

In this subsection, we investigate the relation between the principal eigenvector and the degree vector by their difference and linear correlation coefficient. The degree vector has to be first normalized to quantify its difference with the principal eigenvector. We propose two scalings of the degree vector $\bar{d} = \frac{d}{\sqrt{d^T d}}$ and $\tilde{d} = \frac{\alpha}{\lambda_1} d$, where α is a constant. The corresponding *difference vector* between x_1 and the scaled degree vector is $w = x_1 - \frac{d}{\sqrt{d^T d}}$ and $y = x_1 - \frac{\alpha}{\lambda_1} d$, respectively. The overall difference can be quantified by either the relative difference $u^T w$ (or $u^T y$) or the absolute difference $w^T w$ (or $y^T y$), actually, the sum of the components or the square sum in the difference vector respectively. The first scaling of the degree vector $\bar{d} = \frac{d}{\sqrt{d^T d}}$ aims to obtain the same norm for the degree vector and the principal eigenvector: $\sqrt{\bar{d}^T \bar{d}} = \sqrt{x_1^T x_1} = 1$. The other $\tilde{d} = \frac{\alpha}{\lambda_1} d$ is motivated by $(x_1)_j = \frac{1}{\lambda_1} \sum_{r=1}^N a_{jr} (x_1)_r \leq \frac{d_j}{\lambda_1}$ and the constant α is determined (see Theorem 4.2.1) as the one minimize the absolute difference $y^T y$. Note that both linear scalings of the degree vector will not change the linear correlation coefficient between the principal eigenvector and the degree vector.

Theorem 4.2.1. *The absolute difference $w^T w$ (or $y^T y$) between the principal eigenvector and the degree vector is the smallest ($w^T w = 0$ or $y^T y = 0$) when the spectral radius follows $\lambda_1 = \frac{N_2}{N_1}$, where N_k is the total number of k hop walks between any two nodes which can be the same.*

Proof. The absolute difference

$$w^T w = \left(x_1 - \frac{d}{\sqrt{d^T d}}\right)^T \left(x_1 - \frac{d}{\sqrt{d^T d}}\right) = x_1^T x_1 - 2 \frac{d^T x_1}{\sqrt{d^T d}} + \frac{d^T d}{\left(\sqrt{d^T d}\right)^2} = 2 - 2 \frac{d^T x_1}{\sqrt{d^T d}}. \quad (4.3)$$

Moreover, the generalized form of (4.1) for the k -th largest eigenvalue λ_k and the corresponding eigenvector x_k follow $(x_k)_j = \frac{1}{\lambda_k} \sum_{r=1}^N a_{jr} (x_k)_r = \alpha \frac{d_j}{\lambda_k} - \frac{1}{\lambda_k} \sum_{r=1}^N a_{jr} (\alpha - (x_k)_r)$, we will determine α so that $y_k = x_k - \frac{\alpha}{\lambda_k} d$ has minimum norm. Hence,

$$y_k^T y_k = \left(x_k - \frac{\alpha}{\lambda_k} d\right)^T \left(x_k - \frac{\alpha}{\lambda_k} d\right) = 1 - 2 \frac{\alpha}{\lambda_k} d^T x_k + \frac{\alpha^2}{\lambda_k^2} d^T d, \quad (4.4)$$

is minimized with respect to α if $-\frac{2}{\lambda_k} d^T x_k + 2 \frac{\alpha}{\lambda_k^2} d^T d = 0$ or $\frac{\alpha}{\lambda_k} = \frac{d^T x_k}{d^T d}$. Let $y = y_1$, we obtain

$$y^T y = 1 - \frac{(d^T x_1)^2}{d^T d}, \quad (4.5)$$

using the α derived in the last step. In both Eq. (4.3) and Eq. (4.5), $w^T w = 0$ and $y^T y = 0$ if $d^T x_1 = \sqrt{d^T d}$. In other words, when the principal eigenvector is proportion to degree vector, $w = \mathbf{0}$ (or $y = \mathbf{0}$). Since $Ax_1 = \lambda_1 x_1$, $d^T x_k = \lambda_1 u^T x_1$. The condition $d^T x_1 = \sqrt{d^T d}$ implies

$$\lambda_1 u^T x_1 = \sqrt{d^T d} = \sqrt{N_2}$$

where $N_2 = d^T d$. Since $x_1 = \frac{d}{\sqrt{d^T d}}$, and $u^T d = N_1$, Lemma 4.2.1 follows. \square

Notice that in some approximate mean-field models for virus spreading [73], $\tau_c \sim \frac{N_1}{N_2} = \frac{1}{\lambda_1}$. Furthermore, $w^T w = 0$ (or $y^T y = 0$) is a special case of $u^T y_k = 0$, when $\lambda_1 = \frac{N_2}{N_1}$.

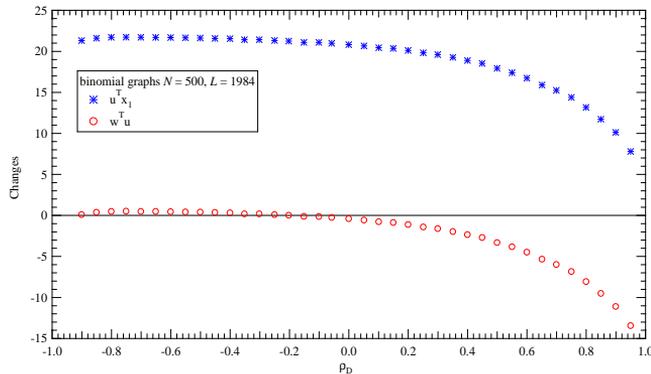


Figure 4.3: Difference between the principal eigenvector and degree vector as a function of the assortativity ρ_D .

The relative difference $w^T u = u^T x_1 - \frac{d^T u}{\sqrt{d^T d}} (y^T u)$ is zero when the absolute difference is zero. We explore the relative difference in general cases by considering the binomial graphs as an example. The sum of the principal eigenvector $u^T x_1$ and the relative difference $w^T u$ as a function of the assortativity are shown in Figure 4.3 to follow exactly the same trend, since the degree of each node, thus, $\frac{d^T u}{\sqrt{d^T d}}$ remains the same when we change the assortativity by degree-preserving rewiring. When the assortativity $\rho_D = 0$, the binomial graphs are actually Erdős-Rényi random graphs, for which $\lambda_1 \simeq \frac{N_2}{N_1}$ when the network size is large [14]. Hence, both the absolute and relative difference are zero when the assortativity is around zero. The sum of the principal eigenvector $u^T x_1$ decreases with the assortativity ρ_D , as explained in Section 4.2.1.

4.2.3 Correlation Between the Principal Eigenvector and the Degree Vector

Recall that so far the best strategy to minimize the spectral radius by links/nodes removal is based on the principal eigenvector. When the correlation $\rho(x_1, d)$ between the principal eigenvector and the degree vector is positively strong, we may use the degree vector instead of the principal eigenvector to determine which links/nodes to remove, which will be further illustrated in Section 4.2.4. Here, we investigate the Pearson correlation coefficient $\rho(x_1, d)$ between the principal eigenvector and the degree vector as a function of ρ_D . Linear scaling of the degree vector will not change the Pearson correlation coefficient. Hence, we consider the original degree vector. When the absolute difference between the principal eigenvector and the scaled degree vector is zero, the principal eigenvector is proportional to the degree vector. In this case, $\rho(x_1, d) = 1$, which seldom occurs in real-world networks. A strong positive correlation, not necessarily to be one, is already interesting with respect to approximate the eigenvector strategy by the corresponding degree vector strategy in minimizing the spectral radius.

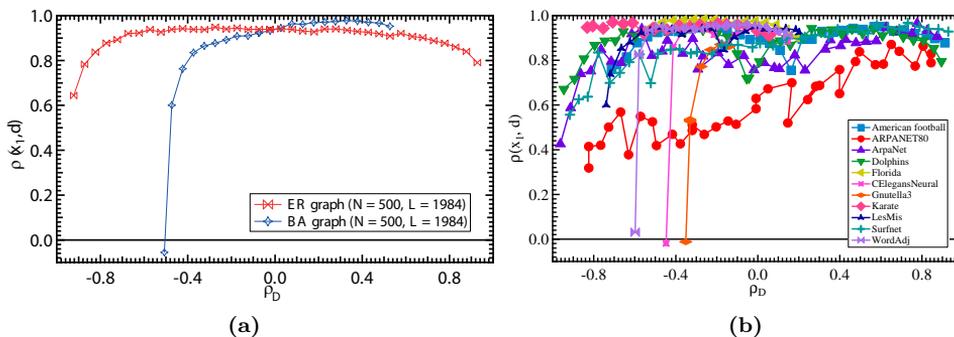


Figure 4.4: Pearson correlation coefficient between the degree vector and the principal eigenvector as a function of the assortativity (a) in both binomial graphs (red marks and line) and power-law graphs (blue marks and line); (b) in network instances derived from real-world networks via degree-preserving rewirings.

Figure 4.4(a) depicts that $\rho(x_1, d)$ is mostly positively strong in the Erdős-Rényi random graphs and B arabasi-Albert graphs. However, $\rho(x_1, d)$ decreases dramatically when the assortativity is decreased, actually around the minimal assortativity. Similarly, we derive networks with different assortativities by applying degree preserving rewiring to each of the 11 real-world networks. As in Figure 4.4(b), We are interested in how $\rho(x_1, d)$ changes with the assortativity ρ_D in real-world networks. Figure 4.4(b) illustrates that, the correlation $\rho(x_1, d)$ creases as the assortativity is decreased, especially around the minimal assortativity, which is the same as observed in network models. In the simulations of both network models and real-world networks, the most evident decrease is observed in networks with a power-law degree distribution such as the C. elegans neural network, the Gnutella 3 network and the WordAdj network.

These observations can be explained similarly as we explain the average/variance of the principal eigenvector versus assortativity in Section 4.2.1. In general, if a node has a large degree, its corresponding principal eigenvector component tends to be large even when the assortativity is zero, due to (4.1). A large positive assortativity implying large (or small) degree nodes tend to connect to other large (or small) degree nodes, further enforces a large degree node to have more likely an even larger principal eigenvector component compared to a small assortativity. Hence, a negative assortativity will weaken the correlation $\rho(x_1, d)$. Note that the correlation coefficient is not necessarily the maximum at the maximal assortativity as shown in Figure 4.4, because here we examine the Pearson correlation coefficient but not the rank correlation.

4.2.4 Application: Degree vs. Principal Eigenvector Strategy in Minimizing the Spectral Radius

In this part, we illustrate the possibility to replace the principal eigenvector strategy by the degree vector in minimizing the spectral radius λ_1 via an example of link removal in power-law networks with different assortativities. As mentioned in [108], so far the best link removal strategy removes the link $l = i \sim j$ with the largest product $(x_1)_i(x_1)_j$ of the

components of the principal eigenvector x_1 . A widely applied strategy to minimize λ_1 by removing m links (1) removes the set of m links with the highest degree product $(d)_i(d)_j$ in the original graph. The corresponding principal eigenvector strategy (2) removes the set of m links with the highest component product in the principal eigenvector of the original graph. We compare these two strategies in removing $m \in [1, 200]$ links in graphs with positive, zero and negative assortativity (see Fig. 4.5) but with the same power-law degree distribution as in Fig. 4.4(a).

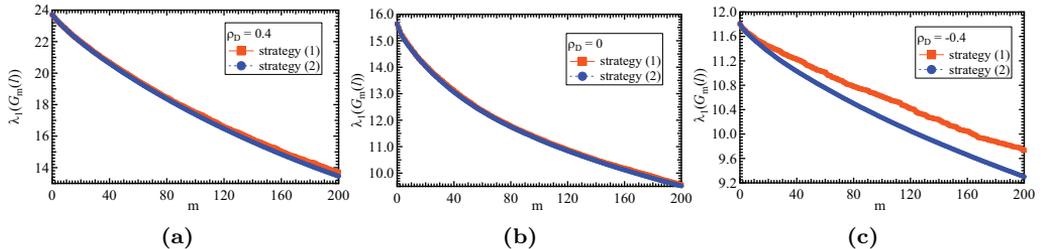


Figure 4.5: Decrease of the spectral radius by successively removing m links in power-law networks. The square and circle dot dash lines show the decrease of the spectral radius by strategies (1) and (2) separately.

Figure 4.5 shows that the decreases of λ_1 by removing links with strategy (1) and (2) are almost same when ρ_D is large. Strategy (2) decreases the spectral radius more thus performs better than strategy (1) when the assortativity is small. When the assortativity is large, the degree vector is positively and strongly correlated with the principal eigenvector. In such a case, the degree vector strategy, the simplest to compute, well approximates the principal eigenvector strategy in minimizing the spectral radius.

4.2.5 Section Conclusions

The principal eigenvector is essential in characterizing the influence of link/node on the spectral radius, whereas its topological meaning is far from well understood. This work, via both theoretical analysis and systematic simulations, contributes to the following aspects: (a) the average $E[x_1]$ (or variance) of the principal eigenvector is shown and explained to decrease (or increase) with the assortativity ρ_D ; (b) the difference between the principal eigenvector and the degree vector is proved to be the smallest, when $\lambda_1 = \frac{N_2}{N_1}$ and (c) we illustrate and explain why the correlation between principal eigenvector and the degree vector decreases as ρ_D is decreased. In general, both a large variance (heterogeneity) in nodal degree and a large degree correlation (homogeneity in connection) contribute to a large average and a small variance of the principal eigenvector and a strong correlation between the degree and the principal eigenvector. As a straightforward application of these finds, we illustrate that when the assortativity is large, we could approximate the well performance principal eigenvector based strategy (to minimize λ_1 by removing links/nodes) by the corresponding degree vector, which is the simplest network property to compute.

4.3 Analyzing the Correlations Between Centrality Metrics

We investigate the correlations between the centrality metrics introduced in Section 2.1.3, in both network models and real-world networks. The network models include the Erdős-Rényi (ER) networks and scale-free (SF) networks. We consider 34 real-world networks, *e.g.*, airline connections, power grids and coauthorship collaborations. The descriptions and properties of these real-world networks are given in Appendix B. We study the correlations between any two centrality metrics using the Pearson correlation coefficient and the centrality similarity.

4.3.1 Pearson Correlation Coefficients Between Centrality Metrics

Here we explore the linear correlation between the centrality metrics using numerical simulations in both ER and SF networks, as well as in real-world networks. The results in Appendix C indicate that strong linear correlations do exist between certain centrality metrics in both ER and SF networks, and that network size has little influence on the correlations. Note that the k -shell index is weakly correlated with all the other centrality metrics. This might be the case because the k -shell indices of all nodes are similar to each other in binomial networks. We note the following seemingly universal relations between the degree masses and three centrality metrics, the principal eigenvector x_1 , the closeness C_n and the betweenness B_n , as

$$\begin{cases} \rho(X_1, D^{(2)}) > \rho(X_1, D^{(1)}) > \rho(X_1, D), \\ \rho(C_n, D^{(1)}) > \rho(C_n, D^{(2)}) > \rho(C_n, D), \\ \rho(B_n, D) > \rho(B_n, D^{(2)}) > \rho(B_n, D^{(1)}), \end{cases}$$

in most real-world networks (see Figs. 4.6a, 4.6b, and 4.6c). The same results can be found in both ER and SF networks (see Appendix C). We theoretically prove the inequality $\rho(X_1, D^{(2)}) > \rho(X_1, D^{(1)}) > \rho(X_1, D)$ in ER networks in Sec. 4.4.

Almost all of the Pearson correlation coefficients $\rho(X_1, D^{(2)})$, $\rho(C_n, D^{(1)})$, and $\rho(B_n, D)$ are large (> 0.95) in both ER and SF networks (see Figs. C.1 and C.2) and are also large (> 0.6) in most real-world networks (see Fig. A.1). The betweenness of a power-law distributed network also follows a power-law distribution [112]. This supports the strong linear correlation between the betweenness B_n and the degree D in SF networks [106].

4.3.2 Centrality Similarities $M_{A,B}(\Upsilon)$ Between Centrality Metrics

Different centrality metrics rank the nodes in different orders within a network. The centrality similarity was proposed in Ref. [113] to quantify the similarity of centrality metrics in ranking nodes

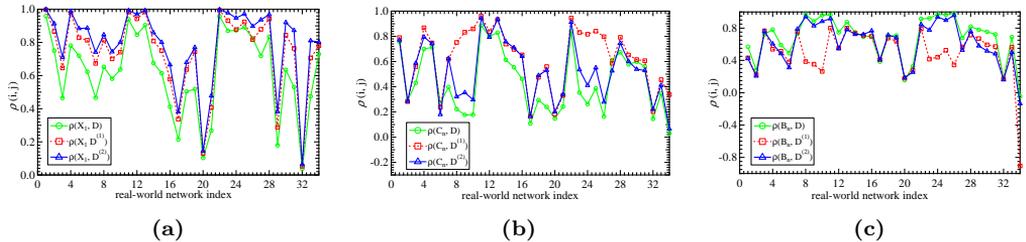


Figure 4.6: Pearson correlation coefficients (a) between the principal eigenvector and the degree masses: $\rho(x_1, D)$ (in circle marks), $\rho(x_1, D^{(1)})$ (in rectangle marks), and $\rho(x_1, D^{(2)})$ (in triangle marks); (b) between the closeness and the degree masses: $\rho(C_n, D)$ (in circle marks), $\rho(C_n, D^{(1)})$ (in rectangle marks), and $\rho(C_n, D^{(2)})$ (in triangle marks); (c) between betweenness and degree masses: $\rho(B_n, D)$ (in circle marks), $\rho(B_n, D^{(1)})$ (in rectangle marks), and $\rho(B_n, D^{(2)})$ (in triangle marks), in 34 real-world networks.

Definition In a graph $G(N, L)$ assume we obtain two node rankings, $[a_{(1)}, a_{(2)}, \dots, a_{(N)}]$ and $[b_{(1)}, b_{(2)}, \dots, b_{(N)}]$, according to centrality metrics A and B , where $a_{(j)}$ or $b_{(j)}$ is the node whose centrality metric A or B is the j -th largest in the networks. The centrality similarity $M_{A,B}(\Upsilon)$ is the percentage of the nodes in $[a_{(1)}, a_{(2)}, \dots, a_{(\Upsilon N)}]$, which are also in $[b_{(1)}, b_{(2)}, \dots, b_{(\Upsilon N)}]$, where $\Upsilon \in [0, 1]$.

The measure $M_{A,B}(\Upsilon)$ gives the percentage of overlapping nodes from the top $100\Upsilon\%$ of nodes, ranked by the centrality metrics A and B , respectively. The range of $M_{A,B}(\Upsilon)$ is between $[0, 1]$. If the $100\Upsilon\%$ of nodes chosen by centrality metric A are not at all in the $100\Upsilon\%$ of nodes chosen by centrality metric B , $M_{A,B}(\Upsilon) = 0$. It means that the most important (top $100\Upsilon\%$) nodes chosen by the two centrality metrics are completely different, *i.e.*, the centrality metrics A and B differ greatly. When all nodes are chosen ($\Upsilon = 1$) there is a full overlap, which indicates that $M_{A,B}(1) = 1$. For a given $\Upsilon < 1$, a larger $M_{A,B}(\Upsilon)$ represents a stronger correlation between the two centrality metrics A and B .

Centrality similarities in network models

We study the centrality similarity $M_{A,B}(\Upsilon)$ between any two centrality metrics² in 10^3 network realizations of ER networks and SF networks with $N = 10^4$ and $\Upsilon = [0.001, 0.01, 0.1]$.

We observe that in both ER and SF networks, the $M_{B_n, D}(\Upsilon)$ is notably larger than the centrality similarity between B_n and any other centrality metric; $M_{C_n, D^{(1)}}(\Upsilon) > M_{C_n, D^{(2)}}(\Upsilon) > M_{C_n, D}(\Upsilon)$; and the centrality similarities $M_{x_1, D^{(1)}}(\Upsilon)$ and $M_{x_1, D^{(2)}}(\Upsilon)$ are both large. In ER networks, $M_{x_1, D^{(2)}}(\Upsilon) > M_{x_1, D^{(1)}}(\Upsilon) > M_{x_1, D}(\Upsilon)$. The k -shell index has low similarity with other metrics in ER networks for the same reason mentioned in Sec. 4.3.1. All these observations agree with what we have found using the Pearson correlation coefficients in Sec. 4.3.1.

²Our study shows that the centrality similarity $M_{A,B}(\Upsilon)$ increases with the increase of Υ in ER networks, but decreases with the increase of Υ in SF networks. Note that this observation holds only for small Υ and, if Υ is around 1, $M_{A,B}(\Upsilon) = 1$ in all networks.

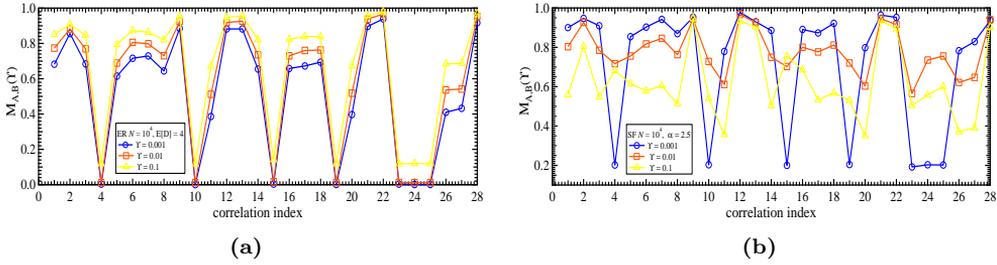


Figure 4.7: Centrality similarities between centrality metrics in network models: (a) for ER networks and (b) for SF networks. The x-axis is the correlation index (see Appendix C).

Centrality similarities in real-world networks

For the 34 real-world networks the percentage Υ should be larger than 3%, since the smallest network only has 35 nodes. We compare the similarity between each centrality metric (e.g., B_n) and all other metrics to determine which metric is the closest to the centrality metric (e.g., B_n). In Fig. 4.8 the height of each bar indicates the number of networks in which $M_{A,B}(\Upsilon)$ is the highest among the centrality similarities between A and all the other centrality metrics. The bar chart shows that the D , $D^{(1)}$, and $D^{(2)}$ are, respectively, most similar to B_n , C_n , and x_1 in most real-world networks, which is consistent with what is observed in the network models. We also observe that either $M_{L_n,D}(\Upsilon)$ or $M_{L_n,B_n}(\Upsilon)$ is the largest among the centrality similarities between L_n and all other metrics in most real-world networks.

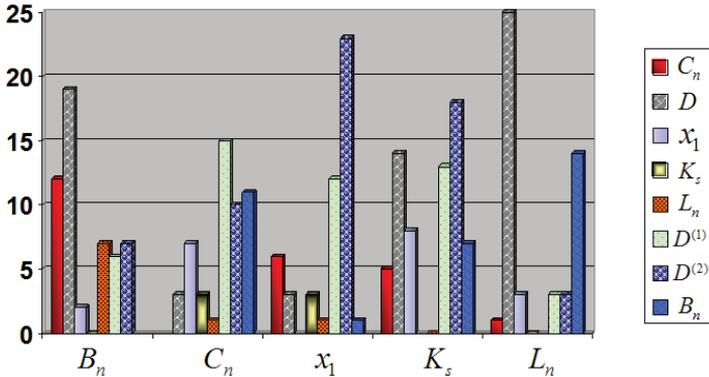


Figure 4.8: Number of networks (among the 34 real-world networks), in which $M_{A,B}(\Upsilon)$ is the highest among the centrality similarities between A and all other centrality metrics, when $\Upsilon = 5\%$. The centrality metric A is given by the x-axis label, and B is reflected by the pattern described in the box on right side. Take the betweenness B_n as an example. The centrality similarities between B_n and all the other metrics are compared with each other to find the largest similarity in each real-world network. For instance, the $M_{B_n,C_n}(\Upsilon)$ is the largest centrality similarity in ‘Electric_s208’ network, so that one is counted into the leftmost bar of B_n (with C_n).

4.4 Theoretical Analysis

The above simulations indicate that the three lowest-order degree masses, with a low computational complexity, are strongly correlated with the betweenness, the closeness, and the components of the principal eigenvector, all of which are complex to compute. We first apply the generating function method [16, 114] to analyze such statistical properties of the degree masses as expectation and variance. We then prove that the high-order ($m \rightarrow \infty$) degree mass is proportional to the principal eigenvector x_1 in any network. Next we prove that when m is small the correlation between degree mass and the principal eigenvector increases with an increase in m , *i.e.*, $\rho(X_1, D^{(2)}) \geq \rho(X_1, D^{(1)}) \geq \rho(X_1, D)$.

4.4.1 Expectation and Variance of the Degree Masses

Lemma 4.4.1. *In an Erdős-Rényi (ER) random network $G_p(N)$, when $N \rightarrow \infty$, the average 1st-order degree mass is*

$$E[D^{(1)}] = N(2p + p^2N), \quad (4.6)$$

and the variance is

$$\text{Var}[D^{(1)}] = N(2p + 4p^2N + p^3N^2). \quad (4.7)$$

The average and the variance of 2nd-order degree mass are

$$E[D^{(2)}] = N(2p + 3p^2N + p^3N^2), \quad (4.8)$$

$$\text{Var}[D^{(2)}] = N(2p + 14p^2N + 17p^3N^2 + 7p^4N^3 + p^5N^4). \quad (4.9)$$

Proof. The generating function for the probability distribution of node degree is defined as

$$\varphi_D(z) = \sum_{k=0}^{N-1} z^k \text{Prob}[D = k],$$

and the generating function of the degree of the node that we arrive at by following a randomly chosen link is

$$\frac{\sum_k k \text{Prob}[D = k] z^k}{\sum_k k \text{Prob}[D = k]} = z \frac{\varphi'_D(z)}{E[D]}, \quad (4.10)$$

where $E[\cdot]$ is the expectation. If we start at a randomly chosen node, the generating function of the degree of a nearest neighbor of this node follows Eq. (4.10). The 1st-order degree mass $D^{(1)}$ of a node equals the degree sum of the node and its neighbors. The generating function has the ‘powers’ property [114], that the distribution of the 1st-order degree mass of a node obtained from one nearest neighbor is generated by

$$\varphi_D(z)^* = z^2 \frac{\varphi'_D(z)}{E[D]},$$

then, the distribution of the total of the 1st-order degree mass over k independent realizations (k nearest neighbors) of the node is generated by k th power of $\varphi_D(z)^*$ as

$$\varphi_{D^{(1)}}(z) = \varphi_D(\varphi_D(z)^*) = \sum_k \text{Prob}[D = k] \left(z^2 \frac{\varphi'_D(z)}{E[D]} \right)^k. \quad (4.11)$$

For ER networks, $E[D] = (N - 1)p$ is the average degree in an ER network $G_p(N)$, and $\varphi_D(z) = (1 - p + pz)^{N-1}$, thus,

$$\varphi_{D^{(1)}}(z) = ((1 - p) + z^2 p (1 - p + pz)^{N-2})^{N-1}, \quad (4.12)$$

In addition, the generating function has the ‘Moments’ property [114], that $E[(D^{(1)})^n] = [(z \frac{d}{dz})^n \varphi_{D^{(1)}}(z)]_{z=1}$. Together with $\text{Var}[D^{(1)}] = E[(D^{(1)})^2] - E[D^{(1)}]^2$, we arrive at the (4.6) and (4.7), when $N \rightarrow \infty$.

Similarly, the distribution of $D^{(2)}$ is generated by $\varphi_D(\varphi_{D^{(1)}}(\varphi_{D^{(1)}}(z)))$. Hence, we obtain the generating function of the 2nd-order degree mass as

$$\varphi_{D^{(2)}}(z) = (1 - p + pz^2(1 - p + pz)^{N-2}(1 - p + pz^2(1 - p + pz)^{N-2})^{N-2})^{N-1},$$

Thus, we can obtain (4.8) and (4.9). \square

4.4.2 Correlation Between the Degree Masses and the Principal Eigenvector

Lemma 4.4.2. *The m th-order degree mass vector $d^{(m)}$ is proportional to the principal eigenvector x_1 in any network with a sufficiently large spectral gap when $m \rightarrow \infty$.*

Proof. The m th-order degree mass vector $d^{(m)}$ is

$$\begin{aligned} d^{(m)} &= \sum_{k=1}^{m+1} (A^k u) = \sum_{k=1}^{m+1} \sum_{j=1}^N \lambda_j^k x_j (x_j^T u) \\ &= \sum_{j=1}^N \left(\lambda_j \frac{\lambda_j^{m+1} - 1}{\lambda_j - 1} \right) (x_j^T u) x_j \\ &= \left(\lambda_1 \frac{\lambda_1^{m+1} - 1}{\lambda_1 - 1} \right) (x_1^T u) x_1 + \sum_{j=2}^N \left(\lambda_j \frac{\lambda_j^{m+1} - 1}{\lambda_j - 1} \right) (x_j^T u) x_j \\ &= \left(\lambda_1 \frac{\lambda_1^{m+1} - 1}{\lambda_1 - 1} \right) (x_1^T u) x_1 \left(1 + O \left(\sum_{j=2}^N \left(\frac{|\lambda_j|}{|\lambda_1|} \right)^m \right) \right). \end{aligned}$$

Literature [14] has proved that $x_1^T u > x_j^T u$ for all $1 < j \leq N$. Accordingly, the term $\sum_{j=2}^N \left(\lambda_j \frac{\lambda_j^{m+1} - 1}{\lambda_j - 1} \right) (x_j^T u) x_j$ is small in the graphs with a large spectral gap $(\lambda_1 - \lambda_2)$.

When m increases, $d^{(m)} \rightarrow \left(\lambda_1 \frac{\lambda_1^{m+1} - 1}{\lambda_1 - 1} \right) (x_1^T u) x_1$. Moreover, when m is large, especially when $m \rightarrow \infty$, $O \left(\sum_{j=2}^N \left(\frac{|\lambda_j|}{|\lambda_1|} \right)^m \right) \rightarrow 0$ in any graph. Thus we find that $d^{(m)}$ tends to be proportional to x_1 when m increases in networks with a large spectral gap, and $d^{(m)} \sim \lambda_1^{(m+1)}(x_1)$ in networks when $m \rightarrow \infty$. \square

Lemma 4.4.3. *In large sparse Erdős-Rényi (ER) networks, $\rho(D^{(2)}, X_1) \geq \rho(D^{(1)}, X_1) \geq \rho(D, X_1)$.*

Proof. The eigenvalue equation $Ax = \lambda x$ leads to $\lambda_1^k x_1 = A^k x_1$, from which we obtain $u^T x_1 \sum_{j=1}^m \lambda_1^j = u^T \left(\sum_{j=1}^m A^j \right) x_1$, where $u^T x_1 = NE[X_1]$ and $u^T \sum_{j=1}^{m+1} A^j = (d^{(m)})^T$. Hence, the relation between the principal eigenvector and the m th-order degree mass vector can be expressed as $E[X_1]N \sum_{j=1}^{m+1} \lambda_1^j = (d^{(m)})^T x_1$, leading to

$$E[D^{(m)} X_1] = E[X_1] \sum_{j=1}^{m+1} \lambda_1^j. \quad (4.13)$$

The Pearson correlation coefficient follows as

$$\rho(D^{(m)}, X_1) = \frac{E[D^{(m)} X_1] - E[D^{(m)}]E[X_1]}{\sqrt{\text{Var}[D^{(m)}]}\sqrt{\text{Var}[X_1]}} = \frac{\left(\sum_{j=1}^{m+1} \lambda_1^j - E[D^{(m)}] \right) E[X_1]}{\sqrt{\text{Var}[D^{(m)}]}\sqrt{\text{Var}[X_1]}}. \quad (4.14)$$

The ratio of the two Pearson correlation coefficients is

$$\frac{\rho(D^{(1)}, X_1)}{\rho(D, X_1)} = \frac{\sqrt{\text{Var}[D]}}{\sqrt{\text{Var}[D^{(1)}]}} \left(1 + \frac{(\lambda_1^2 - E[D^2])}{(\lambda_1 - E[D])} \right) \quad (4.15)$$

For large ER graphs, $E[D] = (N-1)p \rightarrow Np$, $E[D^2] = (N-1)^2 p^2 - (N-1)p^2 + (N-1)p \rightarrow N^2 p^2 - Np^2 + Np$ and $\text{Var}[D] = (N-1)p(1-p) \rightarrow Np(1-p)$. From (4.7), we obtain

$$\frac{\sqrt{\text{Var}[D]}}{\sqrt{\text{Var}[D^{(1)}]}} = \sqrt{\frac{(1-p)}{(E[D] + 2)^2 - 2}} > \frac{1}{E[D] + 2}. \quad (4.16)$$

When $N \rightarrow \infty$ and $Np = \zeta$ (ζ is a constant and independent of N), the spectral radius $\lambda_1 \rightarrow \zeta$, in sparse random graphs [115, 116]. With (4.15) and (4.16), $\rho(D^{(1)}, X_1) \geq \rho(D, X_1)$ is proved.

The ratio of the two Pearson correlation coefficients is

$$\frac{\rho(D^{(2)}, X_1)}{\rho(D^{(1)}, X_1)} = \frac{(\lambda_1 + \lambda_1^2 + \lambda_1^3 - E[D^{(2)}]) \sqrt{\text{Var}[D^{(1)}]}}{(\lambda_1 + \lambda_1^2 - E[D^2] - E[D]) \sqrt{\text{Var}[D^{(2)}]}}$$

with (4.8) and $\lambda_1 \rightarrow Np$, when $N \rightarrow \infty$ we arrive at

$$\frac{(\lambda_1 + \lambda_1^2 + \lambda_1^3 - E[D^{(2)}])}{(\lambda_1 + \lambda_1^2 - E[D^2] - E[D])} = 2E[D] + 1.$$

With (4.7) and (4.9), for large sparse random networks, $\rho(D^{(2)}, X_1) \geq \rho(D^{(1)}, X_1)$ is proved. \square

4.5 Application to the Inflexible Contrarian Opinion (ICO) Model

In this section we apply the studied centrality metrics to select the inflexible contrarians in the inflexible contrarian opinion (ICO) model [76] to help one opinion to compete with another. Both network models and three social networks will be considered.

4.5.1 The ICO Model

The ICO model is a variant of the non-consensus opinion (NCO) model [75]. The ICO and NCO models are both opinion competition models in which two opinions exist and compete with each other. In the NCO model opinions are randomly assigned to all agents (nodes). At time $t = 0$ each agent is assigned opinion A with a probability f and opinion B with a probability $1 - f$. At each subsequent time step each agent adopts the opinion of the majority of its nearest neighbors and itself. When there is a tie, the opinion of the agent does not change. All of the updates are made simultaneously in parallel at each step. The system reaches a state in which the opinions A and B coexist and are stable when f is above a critical threshold f_c .

When the NCO model is in the stable state, the ICO model further selects a fraction p_o of agents with opinion A to be the inflexible contrarians who will hold opinion B , will never change their opinion, but will influence the opinion of other agents. The two opinions then compete with each other according to the update rules of the NCO model. The system will reach a new stable state by following these opinion dynamics.

We use S_1 and S_2 to denote the size of the largest and the second largest clusters of agents with opinion A in the new stable state. A phase transition threshold f_c separates two different phases of the stable state. When $f > f_c$, a giant component of agents with opinion A exists and the coexistence of opinions A and B is stable. When $f \leq f_c$, no giant component of agents with opinion A exists ($S_1 = 0$). The f_c depends on p_o . When $p_o = 0$, the ICO model clearly reduces to the classical NCO model and they have the same critical threshold f_c . When $0 < p_o < p^*$, the threshold f_c of the ICO model increases with p_o , but the size S_1 for the final stable state decreases with p_o . When p is above a certain value p^* , the phase transition no longer occurs, and the giant component of agents with opinion A is completely destroyed ($S_1 = 0$).

4.5.2 Strategies of Selecting Inflexible Contrarians Using Centrality Metrics

The final stable state of the ICO model is affected not only by the percentage p_o , but also by how inflexible contrarian agents are selected. Here we select the inflexible contrarians based on their centrality metrics. Li *et al.* [76] studied the ICO model by choosing the inflexible contrarian agents with opinion A either randomly or according to highest degree. The degree strategy is significantly more effective than the random strategy in reducing the size S_1 of the largest opinion A cluster in the stable state when p_o is the same. Here we want to determine which centrality metric used to pick the inflexible contrarians reduces S_1 most efficiently. We also want to determine whether the S_1 decrease is similar when the inflexible contrarians are chosen based on two strongly correlated (with a large Pearson correlation coefficient or a high centrality similarity) centrality metrics. Here the inflexible contrarians are chosen as nodes with highest (i) betweenness, (ii) degree, (iii) 1st-order degree mass, (iv) 2nd-order degree mass, (v) eigenvector component, (vi) k -shell index, or (vii) leverage or (viii) chosen randomly.

4.5.3 Comparison of Inflexible Contrarian Selection Strategies

We first compare the efficiency in decreasing the size S_1 of the largest opinion A cluster in ER and SF networks when choosing the inflexible contrarians using different centrality metrics. We consider ER networks ($N = 10^4$ or 10^5) with $E[D] = 4$, and SF networks

($N = 10^4$ or 10^5) with $\alpha = 2.5$, and perform all the simulations on 10^3 network realizations.

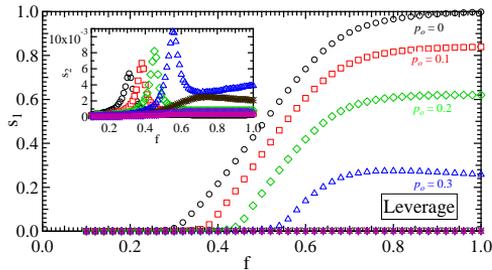


Figure 4.9: An example: the results of leverage strategy. Plot of $s_1 \equiv S_1/N$ as a function of f for different values of p_o for ER networks with $E[D] = 4$ and $N = 10^4$. We denote by S_1 the size of the largest A opinion cluster in the steady-state. Different marks show the results of ICO model with different p_o : $p_o=0$ (\circ), $p_o=0.1$ (\square), $p_o=0.2$ (\diamond), $p_o=0.3$ (\triangle), $p_o=0.4$ ($*$), $p_o=0.5$ (\diamond), $p_o=0.6$ (\boxtimes). The insets plot the $s_2 \equiv S_2/N$, where S_2 is the size of the second largest A opinion cluster, as a function of the f for different values of p_o .

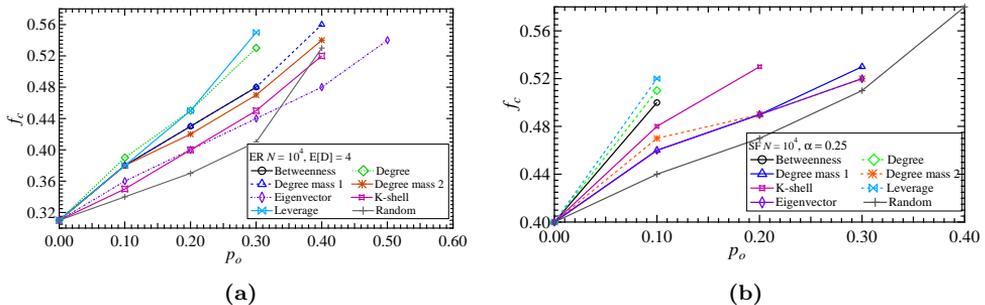


Figure 4.10: Plot of f_c as a function of p_o for strategies 1 to 8: (a) in ER graphs with $N = 10^4$, $E[D] = 4$; (b) in SF graphs with $N = 10^4$, $D_{\min} = 2$, $\alpha = 2.5$.

Figure 4.9 shows a plot of $s_1 = S_1/N$ as a function of f for different values of p_o in ER networks (with $N = 10^4$) using a leverage strategy. The size $s_2 = S_2/N$ shows a sharp peak, a characteristic of a second-order phase transition, in the insets of Fig. 4.9. As p_o increases, f_c shifts to a larger value and the largest cluster becomes significantly smaller. When $p > p^*$, the giant component with opinion A disappears, *i.e.*, $S_1 = 0$. For example, the p^* value for the leverage strategy is between 0.3 and 0.4 (see Fig. 4.9). A small p^* implies that the inflexible contrarians can efficiently destroy the largest opinion A cluster. We can compare the efficiency of the strategies in decreasing S_1 by the value of p^* . When we compare strategies in the ICO model with the same p_o , a larger phase transition f_c for a strategy indicates that the inflexible contrarians chosen using this strategy decreases S_1 more efficiently. Figure 4.10a plots the phase transition f_c as a function of p_o . Note that the efficiency of each strategy is ranked in decreasing order as: Leverage, Degree, Betweenness, 1st-order Degree mass, 2nd-order Degree mass, k -shell

index, Principal Eigenvector, and Random. The same result can be also found in ER and SF networks with $N = 10^5$.

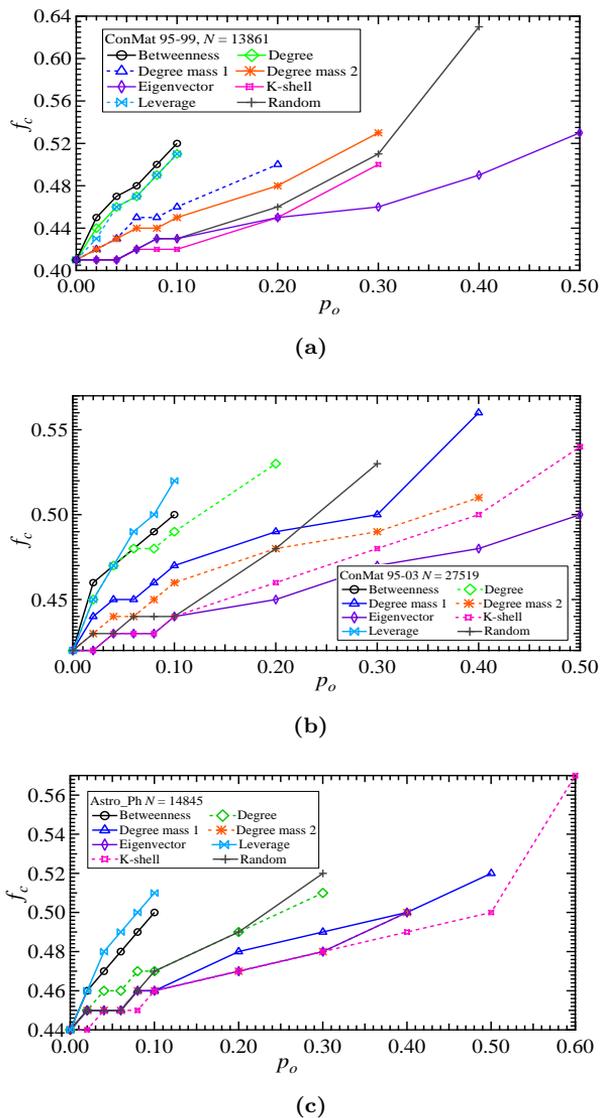


Figure 4.11: Plot of f_c as a function of p_o for strategies in social networks: (a) in network of coauthorships between scientists posting preprints on ConMat E-Print Archives between 1995 to 1999; (b) in network of coauthorships between scientists posting preprints on ConMat E-Print Archives between 1995 to 2003; (c) in network of coauthorships between scientists posting preprints on Astrophysics E-Print Archives between 1995 to 1999.

We find that all strategies are more efficient in SF networks than in ER networks of the same size. We base this on two observations. First, the relative change of f_c with p_o for all strategies in SF networks is larger than it is in ER networks. Second, the p^* for all strategies in SF is much smaller than it is in ER networks. The reason for this may

be that (i) hubs can be readily selected as inflexible contrarians when using centrality metrics in SF networks, and (ii) hubs can strongly influence the opinion of their large number of nearest neighbors.

Figure 4.11 compares these centrality metrics in real-world networks, *i.e.*, the ConMat 95-99 network, the ConMat 95-03 network, and the Astro_Ph network. Note that the inflexible contrarians selected using the leverage L_n , the betweenness B_n , and the degree D are the most efficient in helping opinion B win the competition. The similar behaviors of the three strategies are supported by the large Pearson correlation coefficient $\rho(B_n, D)$ and the large centrality similarities $M_{B_n, D}(\Upsilon)$, $M_{L_n, D}(\Upsilon)$ and $M_{L_n, B_n}(\Upsilon)$.

In both network models and real-world networks, strongly correlated centrality metrics tend to perform similarly. For example, we have discovered both numerically and theoretically that $\rho(D^{(2)}, X_1) \geq \rho(D^{(1)}, X_1)$. Correspondingly, the principal eigenvector x_1 strategy performs closer to the 2nd-order degree mass $D^{(2)}$ than the 1st-order degree mass $D^{(1)}$ in the ICO model.

4.6 Chapter Conclusions

In this chapter we have studied the correlation between widely studied and recently proposed centrality metrics in numerous real-world networks as well as in network models, *i.e.*, as in Erdős-Rényi (ER) random networks and scale-free (SF) networks. A strong correlation between two centrality metrics indicates the possibility of approximating one centrality metric, usually the one with a higher computational complexity, using the other. We find a strong correlation between the degree and the principal eigenvector. When the assortativity is large, we could approximate the well performance principal eigenvector based strategy by the corresponding degree vector. We also study the correlations between the centrality metrics introduced in Sec. 2.1.3 using the Pearson correlation coefficient and the centrality similarity. An important finding is that the degree D , the 1st-order degree mass $D^{(1)}$, and the 2nd-order degree mass $D^{(2)}$ are strongly correlated with the betweenness B_n , the closeness C_n , and the principal eigenvector x_1 , respectively. This observation is partially supported by our analytical proof that $\rho(X_1, D^{(2)}) > \rho(X_1, D^{(1)}) > \rho(X_1, D)$.

We have introduced the degree mass $D^{(m)}$ as a new network centrality metric. The 0th-order degree mass is the degree and the high-order ($m \rightarrow \infty$) degree mass is proportional to the principal eigenvector x_1 . We also find that the influence of network size (the number N of nodes) on the Pearson correlation coefficients is small. In addition, the leverage L_n has high centrality similarities with the degree D and the betweenness B_n . We use these centrality metrics to select the inflexible contrarians in the ICO model to help one opinion to compete with the other. The leverage L_n turns out to be the most efficient strategy in both network models and real-world networks. We also find that strongly correlated metrics perform similarly in the ICO model. This suggests that the metrics with a low computational complexity, such as the degree D and the leverage L_n , could be used to approximate more complex metrics, *e.g.*, the betweenness B_n , to locate important nodes in complex networks. Examples of important nodes would include inflexible contrarians in opinion propagation networks and nodes that should be immunized in disease transmission networks.

CHAPTER 5

Network Metrics for Epidemic Thresholds of SIS Approximations

*“I learned very early the difference between knowing the name
of something and knowing something.”*

Richard P. Feynman, 1918 - 1988

METRICS characterize the functions of a network, in the meanwhile, the functions are usually expressed through the dynamic processes on the network. In this chapter, we introduce the ε -SIS spreading model, which is taken as a benchmark for the comparison between the N-intertwined approximation and the Pastor-Satorras & Vespignani HMF approximation of the SIS model. The N-intertwined approximation, the HMF approximation and the ε -SIS spreading model are compared for different graph types. We focus on the epidemic threshold and the steady-state fraction of infected nodes in networks with different degree distributions. The epidemic thresholds of the approximations can be described by the metrics introduced in Section 2.1. Overall, the N-intertwined approximation is superior to the HMF approximation. The N-intertwined approximation is exactly the same as the HMF approximation in regular graphs. However, for some special graph types, such as the square lattice graph and path graph, the two mean-field approximations are both far away from the ε -SIS spreading model.

5.1 Introduction

As a spreading model we use the susceptible-infected-susceptible [117] (SIS) epidemic process, which is described in Section 2.3.1. A fundamental question in the study of epidemics is, whether a virus will spread through the entire network, or will die out. The answer to this question is the epidemic threshold τ_c , which separates two different phases of the dynamic spreading process on a network: if the effective infection rate τ

is above the threshold, the infection spreads and becomes persistent in time; if $\tau < \tau_c$, the infection dies out exponentially fast. Many authors (see [79, 11, 118, 72, 119, 73, 29, 30]) mention the existence of an epidemic threshold τ_c . Since an exact solution for any network has not been found yet, several approximations of the SIS epidemics have been developed. Here, we focus on the *steady-state* of two *mean-field approximations* of SIS epidemics: the N-intertwined approximation [31, 120] and the Pastor-Satorras & Vespignani approximation [73]. A first order mean-field epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1(A)}$, where $\lambda_1(A)$ is the largest eigenvalue of the adjacency matrix A , was first proposed by Wang *et al.* [30], and rigorously proved by Van Mieghem *et al.* in [31, 120] and later appeared in the physics community [121]. Van Mieghem *et al.* [31] also showed that this mean-field threshold lower bounds the “in reality observed” epidemic threshold, $\tau_c^{(1)} = \frac{1}{\lambda_1(A)} \leq \tau_c$. A more accurate lower bound (the second order mean-field threshold) $\tau_c \geq \tau_c^{(2)} \geq \tau_c^{(1)}$ has been derived in [122]. Pastor-Satorras & Vespignani [73] proposed the heterogeneous mean field (HMF) approximation, whose epidemic threshold [118, 73] is given by $\tau_c^{HMF} = E[D]/E[D^2]$, where D is the degree of a randomly chosen node in G and $\frac{E[D^2]}{E[D]}$ is the degree diversity, which has been introduced in Section 2.1.1.

Here we present a detailed comparison of the two mean-field approximations. Usually, the quality of an approximation is assessed by two criteria: 1) which approximation is closer to the exact SIS model, and 2) which approximation’s epidemic threshold is nearer to the epidemic threshold of the exact SIS model? A direct comparison to the SIS model is, however, not possible, because the steady-state of the exact SIS model in a finite network is, as shown in [31], the overall-healthy state, which is equal to the absorbing state of the SIS Markov chain. The presence of an absorbing state is a major complication in the analysis of the SIS model. The steady-state of both above mean-field approximations corresponds, in fact, to the *meta-stable state* in the SIS model, which is not clearly defined for finite networks [31]. Therefore, we **define** here the meta-stable state of the SIS model via the steady-state of the ε -SIS model for a prescribed value of ε . The ε -SIS process generalizes the SIS model by adding a nodal component to the infection. We assume that each node i can be infected spontaneously. The spontaneous infection process is a Poisson process with rate ε . Hence, besides receiving the infection over links from infected neighbors with rate β , the node i can also itself produce a virus with rate ε . All involved Poisson processes are independent. For $\varepsilon > 0$, the ε -SIS has no absorbing state and Markov theory guarantees a unique steady-state. When $\varepsilon = 0$, the ε -SIS model clearly reduces to the “classical” SIS model. Hence, for small values of $\varepsilon > 0$, the ε -SIS spreading model can be used to approximate the exact SIS model. Here, the ε -SIS spreading model with small value of ε is used as a benchmark to compare the steady-state of the N -intertwined approximation and the HMF approximation on different network types.

This chapter is organized as follows. Section 5.2 overviews the N-intertwined approximation, the Pastor-Satorras & Vespignani HMF approximation and the ε -SIS spreading model in detail. The steady-state of infection in the ε -SIS model and these two approximations are described in Section 5.3. Section 5.4 compares the steady-state fraction of infected nodes in various types of graphs: complete graphs, star graphs, Erdős-Rényi random graphs, small-world graphs, and Barabasi-Albert graphs. An analytic comparison of the epidemic thresholds of the two mean-field approximations is shown in Section 5.5. Conclusions are summarized in Section 5.6.

5.2 Description of the ε -SIS Model and the Mean-field Approximations

5.2.1 The ε -SIS Spreading Model

The ε -SIS spreading model was proposed recently by Hill et al. [123] in their analysis of emotions as a form of infection in a social contact network and earlier in [124] where ε is defined as the driving field conjugate to the density of infected nodes. Here, we will explain the simulation process, but defer to [125] for an analysis of the ε -SIS model.

In our simulations we take a nodal central, event driven approach. An event can either be the curing of a node or the spreading of the infection from one node to another. Events are stored in a timeline as tickets. A ticket contains, besides the time and the event type (spreading or curing), the owner of the ticket. The ticket owner is usually a node, but can also be the system to allow for scheduling of administrative tasks. Tickets are continuously taken from the timeline and passed on to the owner.

If the ticket owner is a node, the ticket either indicates a curing or spreading event. In case of a curing event, the node simply changes its state from infected to healthy; in case of a spreading event, it will spread the infection to the neighbor mentioned in the ticket. If the neighbor was not already infected, it will now become infected and create one or more tickets.

A newly infected node will always create a ticket for its own curing event. According to continuous-time Markov theory (see [16]), the time between infection and curing is exponentially distributed with rate δ and is stored by the node for future reference. An infected node also generates spreading times at which it will spread the infection to its neighbors. The spreading times are again exponentially distributed but now with rate β . If the spreading time does not exceed the node's curing time, a ticket is created for the spreading event. All newly created tickets are stored in the timeline. Finally, the owner of the original ticket generates a new spreading time, which, if not exceeding its own curing time, creates a new spreading ticket for the same neighbor.

If the ticket is not owned by a node, it is a system ticket. System tickets are used to cause the spontaneous infections in nodes. Every node becomes infected spontaneously at a rate ε , but to minimize the number of tickets in the timeline, the system creates one spontaneous infection ticket at the time. The time between spontaneous infection tickets is exponentially distributed with rate $N\varepsilon$. When the system receives a spontaneous infection ticket, it selects a random node and tries to infect it. If the node is already infected, nothing will change, whereas a healthy node will become infected and create the tickets described above.

During the simulation, for each possible number of infected nodes (0 to N) how long the network was in a state with that many nodes infected, is recorded. The average number of infected nodes during the simulation can be determined by multiplying the number of infected nodes by the fraction of time spent in that state, and sum over all the states.

5.2.2 The Pastor-Satorras & Vespignani HMF Approximation

Pastor-Satorras & Vespignani [73] studied the SIS epidemics on networks and proposed the heterogeneous mean-field (HMF) approximation, in which the degree distribution plays

an important role. Highly connected nodes are statistically significant and the strong fluctuations in the degree distribution cannot be neglected. Consider the relative density $\rho_k(t)$ of infected nodes with given degree k , *i.e.*, the probability that a node with k links is infected. The fraction of infected nodes in a network is denoted by ρ . The dynamical mean-field reaction rate equation can be written as

$$\partial_t \rho_k(t) = -\delta \rho_k(t) + \beta k [1 - \rho_k(t)] \Theta(\rho(t)),$$

where $\Theta(\rho(t))$ is the probability that any given link points to an infected node. In steady-state, $y_\infty = \lim_{t \rightarrow \infty} \rho(t)$ is only a function of τ , and as consequence, so is $\Theta(\rho(t))$. By imposing stationarity [$\partial_t \rho_k(t) = 0$], when $t \rightarrow \infty$, the relative density reduces to

$$\rho_k(\tau) = \frac{\tau k \Theta(\tau)}{1 + k\tau \Theta(\tau)}, \quad (5.1)$$

where $\tau = \frac{\beta}{\delta}$ is the effective infection rate and

$$\Theta(\tau) = \frac{1}{E[D]} \sum_{k=1}^{N-1} k \text{Prob}[D = k] \rho_k(\tau). \quad (5.2)$$

Clearly, if $\tau = 0$, then $\Theta(0) = 0$. Substituting (5.1) into (5.2) leads to a self-consistent relation, from which $\Theta(\tau)$ can be determined as

$$\Theta(\tau) = \frac{\tau \Theta(\tau)}{E[D]} \sum_{k=1}^{N-1} \frac{k^2 \text{Prob}[D = k]}{1 + k\tau \Theta(\tau)}. \quad (5.3)$$

Eq. (5.3) has a trivial solution $\Theta(\tau) = 0$. For a nontrivial solution $\Theta(\tau) > 0$ to exist, Eq. (5.3) must satisfy the following condition:

$$\frac{E[D]}{\tau} = \sum_{k=1}^{N-1} \frac{k^2 \text{Prob}[D = k]}{1 + k\tau \Theta(\tau)}. \quad (5.4)$$

Next, we introduce the following expansion,

$$\frac{1}{1 + k\tau \Theta(\tau)} = \sum_{j=0}^{\infty} (-1)^j (k\tau \Theta(\tau))^j$$

valid when $k\tau \Theta(\tau) < 1$ for all k ,

$$\begin{aligned} \frac{E[D]}{\tau} &= \sum_{j=0}^{\infty} (-1)^j \left\{ \sum_{k=1}^{N-1} \text{Prob}[D = k] k^{j+2} \right\} \tau^j \Theta^j(\tau) \\ &= \sum_{j=0}^{\infty} (-1)^j E[D^{j+2}] \tau^j \Theta^j(\tau), \end{aligned}$$

where the latter series converges for $\Theta(\tau) < 1/(D_{max}\tau)$. Since $\tau = 0$ leads to $\Theta(0) = 0$, the non-trivial solution $\Theta(\tau) > 0$ occurs when $\tau > \tau_c^{\text{HMF}} \geq 0$ by the definition of the

epidemic threshold. When $\Theta(\tau)$ is sufficiently small ($\Theta(\tau) < 1/(D_{max}\tau)$) and $\Theta(\tau) > 0$, we can write the above expansion up to first order as

$$\frac{E[D]}{\tau} = E[D^2] - \tau\Theta(\tau)E[D^3] + O(\Theta(\tau)^2), \quad (5.5)$$

in which $\tau\Theta(\tau)E[D^3] > 0$. Hence, when $\tau > \tau_c^{\text{HMF}}$, but $\Theta(\tau)$ is small enough to ignore the second order terms $O(\Theta(\tau)^2)$, with Eq. (5.5) we have

$$\frac{E[D]}{\tau} < E[D^2].$$

It holds that $\tau > \frac{E[D]}{E[D^2]}$, implying that for all $\tau > \tau_c^{\text{HMF}}$. Thus, the epidemic threshold of the HMF approximation is

$$\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]}.$$

The same result was also deduced differently in [126]. For a regular graph [73] with degree r , $E[D^2] = E[D]^2 = r^2$, the epidemic threshold is $\tau_c^{\text{HMF}} = \frac{1}{r} = \frac{1}{\lambda_1}$.

Finally, we can evaluate the fraction $y_\infty(\tau)$ of infected nodes using the relation

$$y_\infty(\tau) = \sum_{k=1}^{N-1} \text{Prob}[D = k] \rho_k(\tau). \quad (5.6)$$

5.2.3 The N-intertwined Approximation

The HMF approximation considers the relative density $\rho_k(t)$ of infected nodes with given degree k during the epidemic process. However, the state of each node is not taken into account. The N-intertwined epidemic approximation [31, 127] is derived by separately observing each node. Every node i at time t in the network is in one of two states: infected, with probability $\text{Prob}[X_i(t) = 1]$ and healthy, with probability $\text{Prob}[X_i(t) = 0]$. Since a node can only be in one of two states, $\text{Prob}[X_i(t) = 0] + \text{Prob}[X_i(t) = 1] = 1$. Since the curing and infected processes are Poisson processes, the whole epidemic process is a Markov process. If we apply Markov theory straightforwardly, the infinitesimal generator $Q_i(t)$ of this two-state continuous Markov chain is

$$Q_i(t) = \begin{bmatrix} -q_{1;i} & q_{1;i} \\ q_{2;i} & -q_{2;i} \end{bmatrix},$$

with $q_{2;i} = \delta$. Markov theory requires that the infinitesimal generator is a matrix whose elements are not random variables. However, this is not the case in our simple approx-

imation: $q_{1;i}(t) = \beta \sum_{k=1}^N a_{ij} 1_{\{X_k(t)=1\}}$. Using a mean-field approximation [31] so that

$E[q_{1;i}] = \beta \sum_{j=1}^N a_{ij} \text{Prob}[X_j(t) = 1]$, the effective infinitesimal generator becomes

$$Q_i(t) = \begin{bmatrix} -E[q_{1;i}] & E[q_{1;i}] \\ \delta & -\delta \end{bmatrix}.$$

Then, in accordance with Markov theory in [16, pp. 182], denoting $v_i(t) = \text{Prob}[X_i(t) = 1]$ and $\text{Prob}[X_i(t) = 0] = 1 - v_i(t)$, the set of nodes obey the differential equations

$$\left\{ \begin{array}{l} \frac{dv_1(t)}{dt} = \beta \sum_{j=1}^N a_{1j} v_j(t) - v_1(t) (\beta \sum_{j=1}^N a_{1j} v_j(t) + \delta) \\ \frac{dv_2(t)}{dt} = \beta \sum_{j=1}^N a_{2j} v_j(t) - v_2(t) (\beta \sum_{j=1}^N a_{2j} v_j(t) + \delta) \\ \vdots \\ \frac{dv_N(t)}{dt} = \beta \sum_{j=1}^N a_{Nj} v_j(t) - v_N(t) (\beta \sum_{j=1}^N a_{Nj} v_j(t) + \delta) \end{array} \right.$$

written in matrix form,

$$\frac{dV(t)}{dt} = \beta AV(t) - \text{diag}(v_i(t))(\beta AV(t) + \delta u), \quad (5.7)$$

where the vector $V(t) = [v_1(t) \ v_2(t) \ \cdots \ v_N(t)]^T$. The average number of infected nodes in G is equal to $y(t) = u^T V(t)$, where u is the all-one vector.

For the N-intertwined approximation, the largest eigenvalue λ_1 of the graph's adjacency matrix rigorously defines the first order epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$. A second order epidemic threshold $\tau_c^{(2)} \geq \tau_c^{(1)}$ is studied in [122] that also presents a different derivation of the N-intertwined equations. The threshold arises as a consequence of the mean-field approximation. A major property, proved in [31] as well as in [122], of the N-intertwined approximation is that $V_i(t) \geq V_i(t)|_{\text{exact}}$. Hence, the N-intertwined approximation upper bounds the SIS epidemics and, consequently, $\tau_c^{(1)} < \tau_c$.

5.3 The Steady-state Infection in the Model and Two Approximations

5.3.1 The ε -SIS Spreading Model

In this chapter, we use the ε -SIS model as a benchmark to compare the two mean-field approximations. Whereas the classical SIS model has an absorbing state, the ε -SIS model does not for $\varepsilon > 0$. The non-zero steady-state of the ε -SIS model is reached as time progresses. We believe that the steady-state fraction of infected nodes in the ε -SIS model is the simplest way to determine the number of infected nodes in the meta-stable state of the SIS model. The meta-stable state of the classical SIS model, although easily recognized, is difficult to define precisely. One approach would be to run many independent instances of the virus spreading process and calculate the average number of infected nodes at sampled points in time and look for a plateau. This will, however, lead to too low an average number of infected nodes as a function of time as for smaller values of the effective spreading rate, many instances of the virus spreading process die out very quickly. These died-out instances have a large impact on the average number of infected nodes as a function of time. Since instances of the virus that die out quickly do not reach a meta-stable state they have to be filtered out, but that would require an assessment of how long a “reasonable” outbreak lasts. Such a reasonable outbreak will

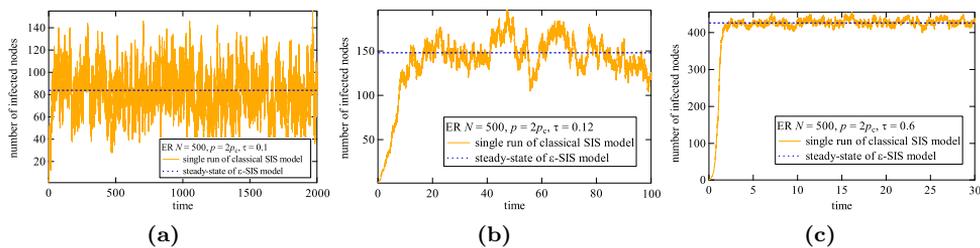


Figure 5.1: Meta-stable state of classical SIS model (solid yellow line) and the steady state of the ε -SIS model (dashed blue line) in ER graphs ($\varepsilon = 10^{-3}$).

be dependent on the effective spreading rate and on the network topology which makes it infeasible as a simulation method.

As the ε -SIS model has a well defined steady-state, the steady-state number of infected nodes can be computed precisely. We start our simulations with no nodes infected and continue to run for a specified warm-up period. After the warm-up period, the measurement period starts during which we record the average number of infected nodes. For all simulations we have taken the warm-up and measurement period to be 10^7 time units and $\varepsilon = 10^{-3}$ time units. We have chosen for a duration of 10^7 time units after careful experimentations. The accuracy of the ε -SIS simulations have been compared to the exact ε -SIS Markov chain (see [125]) for small ($N \leq 10$) networks, where more than 3 digits were accurate for all the considered τ - ranges.

The steady-state number of infected nodes of the ε -SIS model will be close to the average number of infected nodes in the meta-stable state of the SIS model for small values of ε . In Figure 5.1, we show a “reasonable” instance of a virus outbreak together with the steady-state number of infected nodes of the ε -SIS model. These examples illustrate that steady-state average number of infected nodes of the ε -SIS model is precisely the line around which the number of infected nodes in the SIS model varies.

5.3.2 The Pastor-Satorras & Vespignani HMF Approximation

From (5.1) and (5.2), we obtain the set of nonlinear equations

$$\left\{ \begin{array}{l} \frac{\tau \sum_{k=1}^{N-1} k \text{Prob}[D=k] \rho_k}{E[D] + \tau \sum_{k=1}^{N-1} k \text{Prob}[D=k] \rho_k} - \rho_1 = 0 \\ \frac{2\tau \sum_{k=1}^{N-1} k \text{Prob}[D=k] \rho_k}{E[D] + 2\tau \sum_{k=1}^{N-1} k \text{Prob}[D=k] \rho_k} - \rho_2 = 0 \\ \vdots \\ \frac{(N-1)\tau \sum_{k=1}^{N-1} k \text{Prob}[D=k] \rho_k}{E[D] + (N-1)\tau \sum_{k=1}^{N-1} k \text{Prob}[D=k] \rho_k} - \rho_{N-1} = 0 \end{array} \right. \quad (5.8)$$

From the nonlinear set (5.8), the densities $\rho_1, \rho_2, \dots, \rho_{N-1}$ can be calculated, and after using (5.6), we obtain the steady-state fraction $y_\infty(\tau)$ of infected nodes.

5.3.3 The N-intertwined Approximation

The steady-state of the N-intertwined approximation is obtained from (5.7), after letting $t \rightarrow \infty$ and $\lim_{t \rightarrow \infty} \frac{dv_j(t)}{dt} = 0$, as

$$\beta AV(t) - \text{diag}(v_i(t))(\beta AV(t) + \delta u) = 0. \quad (5.9)$$

Written as a nonlinear equation for a single node i , leads to

$$v_{i\infty} = \frac{\beta \sum_{j=1}^N a_{ij} v_{j\infty}}{\beta \sum_{j=1}^N a_{ij} v_{j\infty} + \delta} = 1 - \frac{1}{1 + \tau \sum_{j=1}^N a_{ij} v_{j\infty}}. \quad (5.10)$$

The steady-state fraction $y_\infty(\tau)$ of infected nodes can be calculated using (5.10).

For example, for the complete graphs K_N , with $\tau > \tau_c = \frac{1}{N-1}$, when $t \rightarrow \infty$, $v_{i\infty} = y_\infty$, from which the fraction of infected nodes (5.10) reduces to

$$y_\infty = 1 - \frac{1}{1 + \tau(N-1)y_\infty},$$

or

$$y_\infty = 1 - \frac{1}{(N-1)\tau}, \quad (5.11)$$

which is exactly the same as for the HMF approximation in (5.8) when $\rho_k = \rho_{N-1} = \rho = y_\infty$, as also illustrated in Fig 5.5.

5.3.4 Asymptotics for Large τ

We present the exact steady-state asymptotics of the epidemic for large τ . If τ is sufficiently large, the infection state $v_{j\infty} = \lim_{t \rightarrow \infty} \text{Prob}[X_j(t) = 1]$ of a node j with d_j neighbors tends to be independent of the viral state of its d_j neighbors, because the neighbors are with overwhelming probability infected. Hence, the nodal viral state of node j is not intertwined anymore with that of its neighbors, but independent and is exceedingly well described by a two-state continuous Markov process with infection rate β and curing rate δ , where $v_{j\infty} = \frac{\beta d_j}{\delta + \beta d_j} = \frac{1}{1 + \frac{\delta}{\tau d_j}} = \frac{1}{1 + \frac{s}{d_j}}$, where $s = \frac{\delta}{\tau}$. The derivative for large τ or, equivalently $s \rightarrow 0$, is

$$\left. \frac{dv_{j\infty}(s)}{ds} \right|_{s=0} = -\frac{1}{d_j}.$$

The average steady-state fraction of infected nodes is thus $y_\infty(s) = \frac{1}{N} \sum_{j=1}^N v_{j\infty}(s)$ and has a derivate at $s = 0$ equal to

$$\left. \frac{dy_\infty}{ds} \right|_{s=0} = -\frac{1}{N} \sum_{j=1}^N \frac{1}{d_j} = -E \left[\frac{1}{D} \right],$$

which is precisely equal to that computed in [128] for the N-intertwined mean-field approximation.

For the HMF approximation, we obtain by substituting (5.1) into (5.6) and using the transform $s = \frac{1}{\tau}$,

$$y_{\infty;\text{HMF}}(s) = \sum_{k=1}^{N-1} \frac{\text{Prob}[D = k]}{1 + \frac{s}{k\Theta(s^{-1})}},$$

from which, using $\lim_{s \rightarrow 0} \Theta(s^{-1}) = 1$,

$$\begin{aligned} \left. \frac{dy_{\infty;\text{HMF}}(s)}{ds} \right|_{s=0} &= - \lim_{s \rightarrow 0} \sum_{k=1}^{N-1} \frac{\text{Prob}[D = k]}{\left(1 + \frac{s}{k\Theta(s^{-1})}\right)^2} \frac{1}{k} \left(\frac{1}{\Theta(s^{-1})} - \frac{s}{\Theta^2(s^{-1})} \frac{d\Theta(s^{-1})}{ds} \right) \\ &= - \sum_{k=1}^{N-1} \frac{\text{Prob}[D = k]}{k} = -E \left[\frac{1}{D} \right], \end{aligned}$$

because $\left. \frac{d\Theta(s^{-1})}{ds} \right|_{s=0}$ is finite. Indeed, taking the derivative of the self-consistent relation (5.4)

$$E[D] = \sum_{k=1}^{N-1} \frac{k^2 \text{Prob}[D = k]}{s + k\Theta(s^{-1})},$$

yields

$$0 = \sum_{k=1}^{N-1} \frac{k^2 \text{Prob}[D = k]}{(s + k\Theta(s^{-1}))^2} \left(1 - k \frac{d\Theta(s^{-1})}{ds} \right),$$

or

$$\frac{d\Theta(s^{-1})}{ds} = \frac{\sum_{k=1}^{N-1} \frac{k^2 \text{Prob}[D=k]}{(s+k\Theta(s^{-1}))^2}}{\sum_{k=1}^{N-1} \frac{k^3 \text{Prob}[D=k]}{(s+k\Theta(s^{-1}))^2}},$$

from which $\left. \frac{d\Theta(s^{-1})}{ds} \right|_{s=0} = \frac{1}{\sum_{k=1}^{N-1} k \text{Prob}[D=k]} = \frac{1}{E[D]}$.

Hence, both mean-field approximations return both $\lim_{s \rightarrow 0} y_{\infty}(s)$ and the derivate $\left. \frac{dy_{\infty}(s)}{ds} \right|_{s=0}$ correctly in the large τ -regime.

5.4 Comparison of the Steady-state Fraction of Infected Nodes Versus τ

This section compares the ε -SIS model and two approximations for different graph types. We take the following topologies into account: the bipartite graph, the star graph, the complete graph, the lattice graph, the path graph, the Erdős-Rényi random graph, the Bárábasi-Albert scale-free graph and the small-world graph. These network models are introduced in Section 2.2. The steady-state fraction $y_{\infty}(\tau)$ of infected nodes is calculated for increasing effective spreading rates τ and $\varepsilon = 10^{-3}$. The values of the N-intertwined approximation, the HMF approximation and the simulations of the ε -SIS spreading model are shown in purple, blue and pink lines respectively. The different markers indicate the size of the graphs, *e.g.* circles in Figure 5.2 indicate the results for graphs with $N = 10$ nodes.

5.4.1 Complete Bipartite Graphs

Here we consider complete bipartite graphs K_{M_1, M_2} with $M_1 = N/4$ and $M_2 = 3N/4$. The steady-state fraction $y_\infty(\tau)$ of infected nodes as a function of τ are computed in bipartite graphs with $N = 10, 20, 40, 80, 160$ and 320 nodes. Figure 5.2 shows that the epidemic thresholds for the HMF approximation and the N-intertwined approximation are close to that of the ε -SIS spreading model ($\varepsilon = 10^{-3}$) in complete bipartite graphs. Since, $\tau_c^{(1)}$ of the N-intertwined approximation is nearer to τ_c than τ_c^{HMF} of the HMF approximation, $\tau_c^{(1)}$ provides the better epidemic prediction for the SIS model in the complete bipartite graph K_{M_1, M_2} . Moreover, in [122] it is proved that $\tau_c \geq \tau_c^{(2)} \geq \tau_c^{(1)}$, which means that the second order N-intertwined approximation is closest to the ε -SIS spreading model, and therefore the best in bipartite graphs.

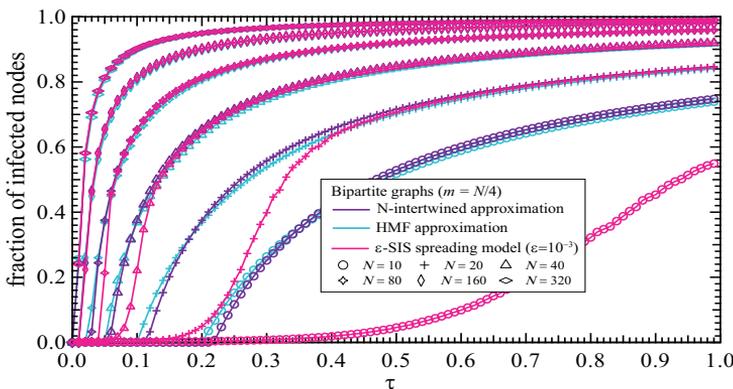


Figure 5.2: Comparison in bipartite Networks

Three interesting results can be observed by zooming in on Figure 5.2 as shown in Figure 5.3. First, the N-intertwined approximation is an upper bound of the ε -SIS spreading model. Second, the difference between the N-intertwined approximation and the ε -SIS spreading model decreases with N . We observe that the N-intertwined approximation almost overlays the ε -SIS spreading model, when $N = 320$. Third, the HMF approximation is lower than the ε -SIS spreading model, showing that the HMF approximation is not upper bounding the SIS model.

5.4.2 Star Graphs

The epidemic threshold for the first order N-intertwined approximation equals $\tau_c^{(1)} = \frac{1}{\lambda_1}$. For any connected graph, the spectral radius is bounded [14] from above by $\lambda_1 \leq \sqrt{2L - N + 1}$, and equality is reached for the complete graph K_N , and the star $K_{1, N-1}$. As a star graph contains $L = N - 1$ links, we obtain

$$\tau_c^{(1)} = \frac{1}{\sqrt{2L - N + 1}} = \frac{1}{\sqrt{N - 1}}. \quad (5.12)$$

The second-order mean-field threshold for the star was estimated in [122] to be $\tau_c^{(2)} = \frac{1}{\sqrt{0.53N - 1.3}}$, while exact computations indicate that $\tau_c = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{2} \log N + \log \log N + O(1)}$ for large N .

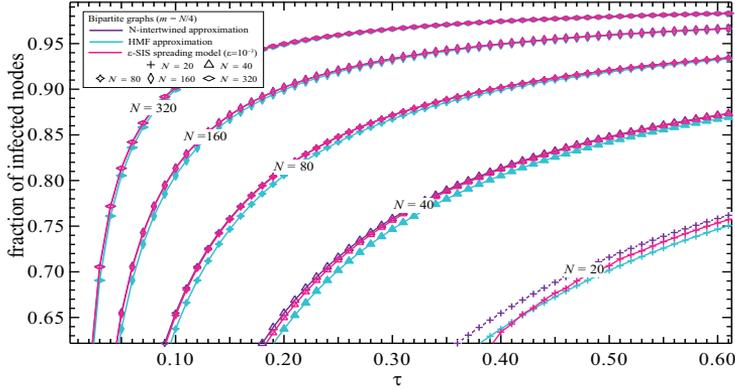


Figure 5.3: Zoom in the comparison in the bipartite graphs

Recall that the epidemic threshold of the HMF approximation is given by $\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]}$. For star graphs it holds that $E[D^2] = \frac{N^2 - N}{N}$ and $E[D] = \frac{2(N-1)}{N}$, so the HMF threshold reduces to

$$\tau_c^{\text{HMF}} = \frac{2}{N}. \quad (5.13)$$

Equalities (5.12) and (5.13) indicate that, for $N > 2$, the epidemic threshold of the N-intertwined approximation is always larger than that of the HMF approximation in star graphs. Figure 5.4 shows the superiority of the N-intertwined approximation, especially when N is large. Nevertheless, the two epidemic thresholds are both quite far from the threshold of the ε -SIS spreading model ($\varepsilon = 10^{-3}$) in star graphs.

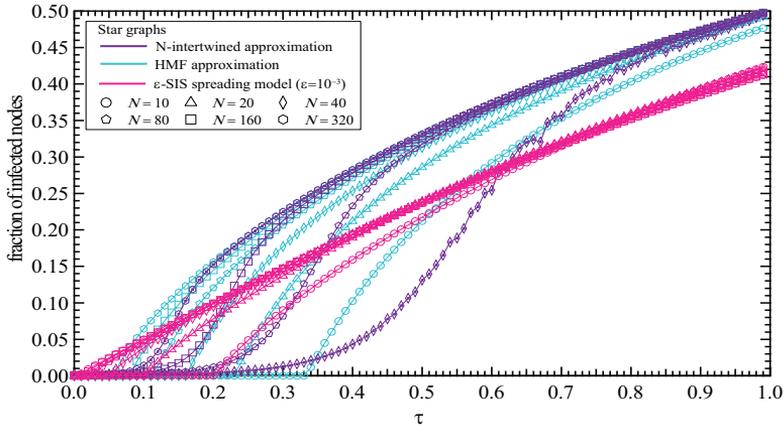


Figure 5.4: Comparison in star graphs

5.4.3 Complete Graphs

For a complete graph $\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]} = \frac{N-1}{N(N-1)^2/N} = \frac{1}{N-1}$, at the same time $\lambda_1 = N-1$. Hence, the epidemic threshold of the N-intertwined approximation $\tau_c^{(1)} = \frac{1}{\lambda_1}$ is equal

to the threshold of HMF approximation $\tau_c^{HMF} = \frac{E[D]}{E[D^2]}$. For K_N , both approximations are very close to the ε -SIS spreading model ($\varepsilon = 10^{-3}$) (see Figure 5.5). This is to be expected, since the mean-field approximation in the N-intertwined approximation is best for dense graphs, as explained in [31]. Moreover, for K_N , the steady-state equations (see Sections 5.3.3 and 5.3.2) in the N-intertwined and HMF approximation are the same. The steady-state fraction $y_\infty(\tau)$ of infected nodes as a function of τ has been deduced in (5.11).

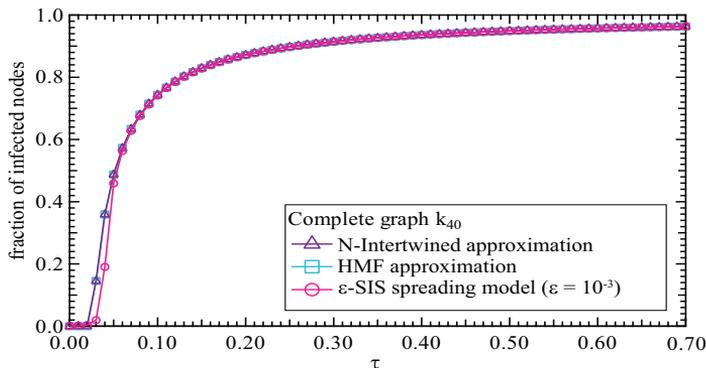


Figure 5.5: Comparison in complete graphs

5.4.4 Square Lattice Graphs

For square lattice graphs, the equations of N-intertwined approximation and the HMF approximation are almost the same, as verified from the simulations of the two approximations. Our simulations (see Figure 5.6) show that the epidemic threshold of the ε -SIS spreading model ($\varepsilon = 10^{-3}$) decreases with the size N of the network. The HMF approximation performs a bit better than the N-intertwined approximation in approaching the ε -SIS spreading model in lattice graphs. The simulation illustrates that both the N-intertwined approximation and the HMF approximation do not predict the epidemic threshold for epidemic processes in lattices. We remark that, in the related process of percolation, the critical probability [61, 62, 63] on the square lattice is equal to $1/2$.

5.4.5 Path Graphs

As shown in Figure 5.7, the steady-state fraction $y_\infty(\tau)$ of infected nodes of the N-intertwined approximation and the HMF approximation are far from that of the ε -SIS spreading model ($\varepsilon = 10^{-3}$). The epidemic thresholds of the N-intertwined approximation and the HMF approximation are both near 0.5, since the average degree of the path graph is 2, ignoring boundary nodes. However, the steady-state fraction $y_\infty(\tau)$ of infected nodes of the ε -SIS spreading model increases very slowly with τ between $0 \leq \tau \leq 1$, and seems to always be around 0 in the range of network sizes that we considered.

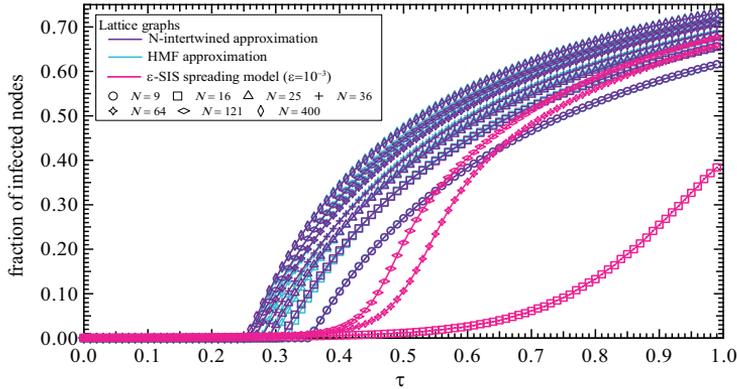


Figure 5.6: Comparison in lattice graphs

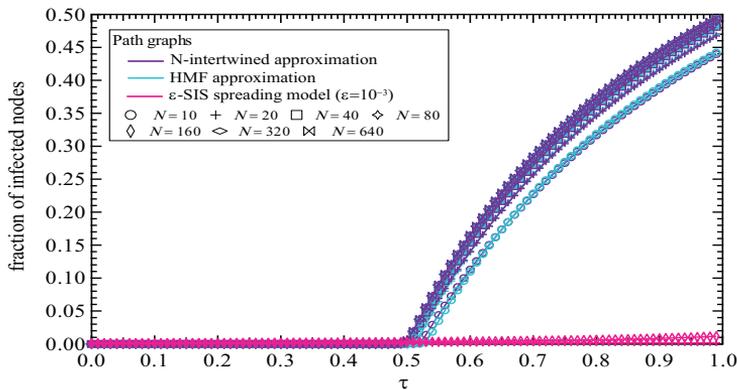


Figure 5.7: Comparison in path graphs

5.4.6 Erdős-Rényi Random Graphs

In this section we investigate the thresholds in Erdős-Rényi (ER) random graphs, which have a binomial degree distribution [1]. An Erdős-Rényi random graph is connected with high probability, if $p > p_c \approx \frac{\ln N}{N}$ for large N , where p_c is the disconnectivity threshold. All the graphs in the simulations are generated with $p = 2p_c$, and checked for connectivity. Figure 5.8 shows that the steady-state fraction $y_\infty(\tau)$ of infected nodes of the N-intertwined approximation and the HMF approximation for ER graphs $N = 10, 20, 40$ and 80 , are extremely close. However, they both differ from the epidemic threshold of the ε -SIS spreading model, especially when N is small. When N is large, the two approximations are close to the ε -SIS spreading model ($\varepsilon = 10^{-3}$) (see Figure 5.9).

5.4.7 Bárábasi-Albert Scale-free Graphs

The Bárábasi-Albert (BA) graph [7] is a characteristic model for complex networks because of its power-law degree distribution. Power-law degree distributions are widely, though approximately, observed in real-world complex networks. The steady-state fraction of infected nodes as a function of the effective spreading rate $y_\infty(\tau)$ is computed in a BA graph with $N = 1000$ and $m = 4$ and shown in Figure 5.10. The N-intertwined

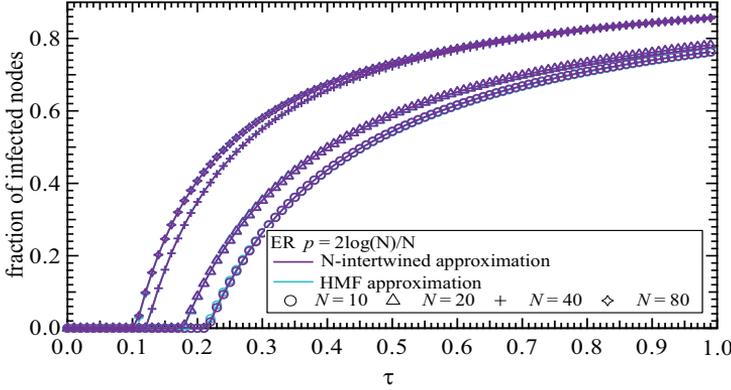


Figure 5.8: Comparison in Erdős-Rényi random graphs

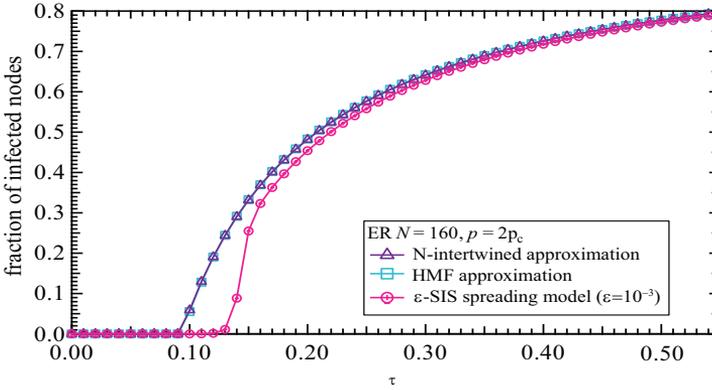


Figure 5.9: Comparison among the N -intertwined approximation, Pastor-Satorras approximation and the ε -SIS model in ER network ($N = 160$).

approximation is close to the HMF approximation, but a little superior. This is to be expected, since the N -intertwined approximation is better than the HMF approximation in star graphs as explained in Section 5.4.2, and the BA model can be regarded as a set of hubs with star graph features.

5.4.8 Watts-Strogatz Small-world Graphs

Watts-Strogatz small-world (WS) graphs [6] have two main properties: a small average hopcount $E[H]$, similar to Erdős-Rényi random graphs, and a high clustering coefficient C_G , similar to a ring lattice. The structural properties of small-world graphs have been found in various real-world networks, including social networks [65], neural networks [66] and biological oscillators [67]. In this section, the WS graphs are generated with $N = 40$ and 80 , $k_s = 6$ and $p = 0.1$ and 1 . In Figure 5.11 the steady-state fraction $y_\infty(\tau)$ of infected nodes, as predicted by the two approximations are shown together with the ε -SIS simulations. The N -intertwined approximation and the HMF approximation are quite close to each other, but far away from the ε -SIS spreading model. The $\tau_c^{(1)} = \frac{1}{\lambda_1}$ and the $\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]}$ in small-world graphs are near to each other no matter what N

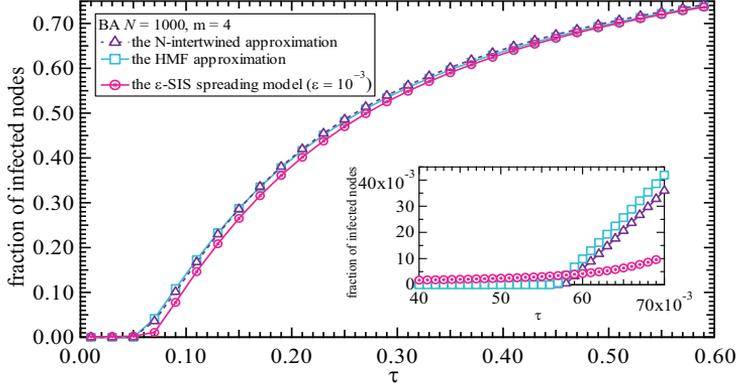


Figure 5.10: Comparison in Barabasi-Albert Scale-free networks

and p are. This can be explained by observing that most nodes have the same degree in WS graphs, justifying the approximation of $E[D^2]$ by $E[D]^2$ and $\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]}$ by $\frac{1}{E[D]}$. Another consequence of the similar node degrees in WS graphs is that $E[D]$ is close to D_{\max} . Since λ_1 is bounded from below and above as $E[D] \leq \lambda_1 \leq D_{\max}$ [14, art. 43, pp. 46 and art. 48, pp.52], we can approximate λ_1 by $E[D]$, and $\tau_c^{(1)}$ by $\frac{1}{E[D]}$, just as τ_c^{HMF} .

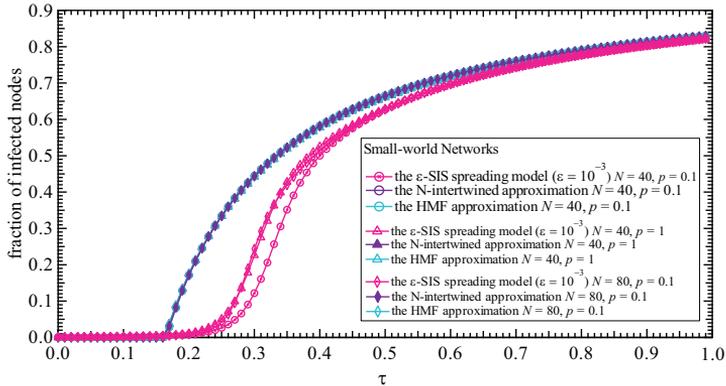


Figure 5.11: Comparison in WS Small-world networks

5.5 Analytic Comparison of the Epidemic Thresholds $\tau_c^{(1)}$ and τ_c^{HMF}

In this paragraph, we analyse the relation between the first order epidemic threshold of N-intertwined approximation $\tau_c^{(1)} = \frac{1}{\lambda_1}$ and the epidemic threshold of the HMF approximation $\tau_c^{\text{HMF}} = \frac{E[D]}{E[D^2]}$. From the comparison in Section 5.4, we find that the relation between the two epidemic thresholds strongly depends on the graph types. The two epidemic thresholds are equal to each other in regular graphs where each node has degree r increasing with N . Indeed, since $\lambda_1 = E[D] = r$ (see [14, art. 43, pp. 46]), and $\tau_c^{\text{HMF}} = \frac{1}{r}$, we find that $\tau_c^{(1)} = \tau_c^{\text{HMF}}$. There are graphs for which $\tau_c^{(1)} < \tau_c^{\text{HMF}}$, while in most cases,

our simulations in Figures 5.2, 5.4, 5.8 and 5.10 demonstrate that $\tau_c^{(1)} > \tau_c^{\text{HMF}}$.

Cases $\tau_c^{(1)} < \tau_c^{\text{HMF}}$: The epidemic threshold τ_c^{HMF} is larger than the first order threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$, when the assortativity ρ_D is zero. In [90], we have proved that $\lambda_1 \geq \frac{N_2}{N_1} = \frac{E[D^2]}{E[D]} = \frac{1}{\tau_c^{\text{HMF}}}$, when $\rho_D = 0$ (see Eq. (3.9)).

Cases $\tau_c^{(1)} > \tau_c^{\text{HMF}}$: Newman [18] pointed out that the assortativity ρ_D of the ER graph and the BA graph is zero when N is large. However, in most ER and BA graphs with finite size, the assortativity is only *approximately* zero. Our simulations in Figures 5.8 and 5.10 show that $\tau_c^{\text{HMF}} \leq \tau_c^{(1)}$ in ER and BA graphs, demonstrating that the precise $\rho_D = 0$ condition in (2.3) that led to $N_1 N_3 = N_2^2$ is not valid. Moreover, we have already proved $\tau_c^{\text{HMF}} \leq \tau_c^{(1)}$ in star graphs (see Sec. 5.4.2).

It would be interesting to find all or the most prominent graph classes in which $\tau_c^{(1)} > \tau_c^{\text{HMF}}$ and in which $\tau_c^{(1)} < \tau_c^{\text{HMF}}$.

5.6 Chapter Conclusions

Many approximations of the SIS model have been proposed to understand SIS epidemics. In this chapter, we studied which mean-field approximation, the N-intertwined or the HMF, is better in approaching the SIS epidemic model. A direct comparison to the SIS model is, however, not possible, because the steady-state of the exact SIS model in a finite network is the overall-healthy state. Although an infection in the SIS model will eventually die out, for high enough effective spreading rates the fraction of infected nodes as a function of time is meta-stable. We proposed to define the number of infected nodes in the meta-stable state of the SIS model via the number of infected nodes in the steady-state of the ε -SIS model for a prescribed small value of ε . From the comparison between the N-intertwined and HMF approximations with the ε -SIS spreading model, we conclude that, overall, the N-intertwined approximation is better than the HMF approximation, except for square lattice graphs and path graphs. We have seen that the N-intertwined approximation can approach the ε -SIS epidemic model well in most graph types. The simulations show that the N-intertwined approximation almost overlaps with the ε -SIS spreading model, when the size of network is large enough. While the HMF approximation is better than the N-intertwined approximation in the square lattice and path graphs, the difference between the two is small. Moreover, they are both far away from the ε -SIS spreading model. We also showed that the N-intertwined approximation and the HMF approximation are exactly the same in regular graphs with the degree of nodes increasing with N , such as complete graphs, and are similar in small-world graphs. In addition to our simulation results, we showed analytically the conditions under which the epidemic threshold of the N-intertwined approximation is larger than, smaller than or equal to that of the HMF approximation.

Part II

NETWORK DESIGN

CHAPTER 6

Increasing the Epidemic Threshold by Link or Node Removals

“What goes up must come down.”

Isaac Newton, 1642 - 1727

CHAPTER 5 has shown that the largest eigenvalue $\lambda_1(A)$ of the adjacency matrix A plays an important role in dynamic processes in networks, *e.g.* virus spreading [31]. In a SIS type of network infection, the steady-state¹ infection of the network is determined by a phase transition at the epidemic threshold $\tau_c = \frac{1}{\lambda_1(A)}$: when the effective infection rate $\tau > \tau_c$, the network is infected, whereas below τ_c , the network is virus-free. Beside virus spread, the same type of phase-transition threshold [44] in the coupling strength $g_c \sim \frac{1}{\lambda_1(A)}$ occurs in a network of coupled oscillators. Motivated by a $\frac{1}{\lambda_1(A)}$ threshold separating two different phases of a dynamic process on a network, we want to design networks in order to enlarge the network's epidemic threshold τ_c , or, equivalently, to lower $\lambda_1(A)$. In this chapter, we concentrate mainly on the following problems: removing m links from a graph G and removing m nodes from a graph G .

The minimization of the spectral radius by removing m links is shown to be an NP-complete problem, which suggests to consider heuristic strategies. Several greedy strategies are compared and several bounds on the decrease of the spectral radius are derived in [108]. The strategy that removes that link $l = i \sim j$ with largest product $(x_1)_i (x_1)_j$ of the components of the eigenvector x_1 belonging to the largest adjacency eigenvalue is shown to be superior to other strategies in most cases. Furthermore, a scaling law where the decrease in spectral radius is inversely proportional to the number of nodes

¹In the exact SIS model, the steady-state is the healthy state, which is the only absorbing state in the Markov process. However, in networks of realistic size N , this steady-state is only reached after an unrealistically long time. The steady-state in the N -intertwined virus spread model refers to the meta-stable state, which is reached exponentially rapidly and which reflects real epidemics more closely.

N in the graph is deduced. Another sublinear scaling law of the decrease in spectral radius versus the number m of removed links is conjectured.

We also present a new type of lower bound for the spectral radius of a graph in which m nodes are removed [129]. Moreover, we present a lower bound and an upper bound for the components of the principal eigenvector. As a corollary, Cioabă's theorem [130], which states that the maximum normalized principal eigenvector component in any graph never exceeds $\frac{1}{\sqrt{2}}$ (with equality for the star), appears as a special case of our more general result.

6.1 Introduction

We are searching for a strategy so that, after removing m links (or nodes), λ_1 is minimal. Earlier, Restrepo *et al.* [131] have initiated an instance of this problem: "How does λ_1 decrease when links are removed?" They introduced a new graph metric, called the dynamical importance $I_x = \frac{\lambda_1(A) - \lambda_1(A \setminus \{x\})}{\lambda_1(A)}$, where x either denotes the removal of a link $x = l$ or of a node $x = n$. The dynamical importance was further investigated by Milanese *et al.* [132]. Both Restrepo *et al.* and Milanese *et al.* have approached the problem by using perturbation theory. However, they did not consider optimality of their removal strategy.

We first introduce the notations used in this Chapter. We consider a graph $G = (\mathcal{N}, \mathcal{L})$, where \mathcal{N} is the set of nodes and \mathcal{L} is the set of links. The number of nodes is denoted by $N = |\mathcal{N}|$ and the number of links is represented by $L = |\mathcal{L}|$. Let x_1 be the eigenvector of A belonging to $\lambda_1(A)$ in the original graph G and normalized such that $x_1^T x_1 = 1$. Let \mathcal{L}_m (or \mathcal{N}_m) denote the set of the m links (or nodes) that are removed from G , and $G_m(\mathcal{L}) = G \setminus \mathcal{L}_m$ (or $G_m(\mathcal{N}) = G \setminus \mathcal{N}_m$) is the resulting graph after the removal of m links (or nodes) from G . We denote the adjacency matrix of $G_m(\mathcal{L})$ (or $G_m(\mathcal{N})$) by $A_m(\mathcal{L})$ (or $A_m(\mathcal{N})$), which is still a symmetric matrix. Similarly, let w_1 be the normalized eigenvector (as in [108]) of $A_m(\mathcal{L})$ (or $A_m(\mathcal{N})$) corresponding to $\lambda_1(A_m(\mathcal{L}))$ (or $\lambda_1(A_m(\mathcal{N}))$) in the graph $G_m(\mathcal{L})$ (or $G_m(\mathcal{N})$) (such that $w_1^T w_1 = 1$). By the Perron-Frobenius theorem [14], all components of x_1 and w_1 are non-negative (positive if the corresponding graph is connected).

In this chapter, we complement their study by first showing in Section 6.2 that the *Link Spectral Radius Minimization (LSRM) problem and the Nodal Spectral Radius Minimization (NSRM) problem*, defined in Problem 1 and Problem 3, are NP-hard, which means in practice, that an optimal solution in a large network cannot be computed and that good approximate algorithms or heuristics need to be devised. The NP-completeness of LSRM and NSRM is demonstrated by reducing the problem to an equivalent problem, namely finding a Hamiltonian path in a graph, that is known to be NP-complete [133]. Since LSRM and NSRM are NP-complete, we cannot hope to find exact analytic formulae for the decrease in the spectral radius. However, in Section 6.3.1, we provide a general analytic description, bounds, several lemmas and we study the effect of node and link removal on closed walks and the influence of assortativity on the spectral radius. This developed theory directs us to find good heuristics. Section 6.3.2 proposes eight different strategies (or heuristics) for removing one link in a network and these strategies are benchmarked with the optimal strategy via extensive simulations. The removal of the link l between nodes i and j with highest product $(x_1)_i (x_1)_j$ of the eigenvector compon-

ents belonging to the largest eigenvalue $\lambda_1(A)$ of the adjacency A is demonstrated to be the best heuristic. However, it is not always the best heuristic when more than one link is removed as illustrated in Figures 6.1-6.3. The scaling law (6.12) for removing one link in Section 6.3.3 demonstrates, presumably for all graphs, that a decrease in λ_1 is inversely proportional to the size N of the graph. Hence, small graphs show the effect of link removals on λ_1 more clearly than large graphs. The scaling law (6.9) is much less accurately known, but indicates a sublinear decrease in λ_1 with the number m of removed links. We also claim that the optimal way to remove m links is to make the resulting graph as regular as possible, because a regular graph has the lowest spectral radius among all graphs with N nodes and L links. Section 6.4 studies the bounds for the decrease of the spectral radius by node removals.

Another type of strategy to prevent the outbreak of a virus is to quarantine infected nodes. Omic *et al.* [134] have studied immunization via modularity partitioning, where inter-community links are removed such that intra-community communication is preserved. Taylor and Restrepo [135] investigated the effect of adding a subgraph to a network on its largest adjacency eigenvalue λ_1 . Inspired by network synchronization, Watanabe and Masuda [50] have investigated a similar problem as the NSRM but with a different object function: remove nodes in a graph to maximize the second smallest eigenvalue of the Laplacian of the graph, also coined the algebraic connectivity [14]. They have presented several strategies comparable to ours, and also found that the eigenvector strategy performed overall the best. Related to [50], but based on a weighted, asymmetric Laplacian of a graph, Nishikawaa and Mottera [136] point to the non-trivial effect of link removals on network synchronization.

6.2 The Spectral Radius Minimization Problem is NP-hard

In this section, we prove that optimally decreasing the largest adjacency eigenvalue (the spectral radius) of a graph by a fixed number of link or node removals is NP-hard. It is widely believed that NP-hard problems cannot be solved exactly in a time complexity that is upper bounded by a polynomial function of the relevant input parameters (N and L).

6.2.1 The Link Spectral Radius Minimization Problem is NP-hard

Let us first formulate the *Link Spectral Radius Minimization (LSRM) problem* precisely:

Problem 1 (Link Spectral Radius Minimization (LSRM) problem). *Given a graph $G(\mathcal{N}, \mathcal{L})$ with N nodes and L links, spectral radius $\lambda_1(G)$, and an integer number $m < L$. Which m links from the graph G need to be removed, such that the spectral radius of the reduced graph G_m of $L - m$ links has the smallest spectral radius out of all possible graphs that can be obtained from G by removing m links?*

Theorem 6.2.1. *The LSRM problem is NP-hard.*

To prove this theorem, we rely on the following lemmas, but first we need the definition of a path P_h with h hops or links. A path P_h with h hops starting from a node n_0 and ending at node n_h is defined as $P_h = n_0 \sim n_1 \sim n_2 \sim \dots \sim n_{h-1} \sim n_h$, where each link $n_i \sim n_j$ between nodes n_i and n_j as well as each node n_i occurs once in the sequence

defining the path P_h , in contrast to a walk $W_h = n_0 \sim n_1 \sim n_2 \sim \dots \sim n_{h-1} \sim n_h$ with h hops, where a node n_i can appear more than once.

Lemma 6.2.2. *The path P_{N-1} visiting N nodes has a strictly smaller spectral radius than all other connected graphs with N nodes. Furthermore, $\lambda_1(P_{N-1}) = 2 \cos\left(\frac{\pi}{N+1}\right)$.*

Proof: [137, p. 21][14, p. 125] □

Lemma 6.2.3. *The eigenvalues of a disconnected graph are composed of the eigenvalues (including multiplicities) of its connected components.*

Proof: [14, p. 73-74] □

Lemma 6.2.4. *Among all possible graphs of N nodes and $N - 1$ links, the path P_{N-1} visiting N nodes has the smallest spectral radius.*

Proof: A connected graph of N nodes and $N - 1$ links is a tree, of which the path is a special case. According to Lemma 6.2.2 the path has a spectral radius strictly smaller than 2, which is the smallest spectral radius possible in connected graphs. Hence, we need to demonstrate that the Lemma also holds for disconnected graphs. For ease of presentation, we assume that the disconnected graph consists of two connected components A_1 and A_2 : A_1 of x nodes and A_2 of $N - x$ nodes. Our arguments also apply to multiple connected components. Now, A_1 contains at least $x - 1$ links, otherwise it is not a connected component, and A_2 contains at least $N - x - 1$ links. Since the sum of these links equals $N - 2$, either A_1 or A_2 must contain one extra link, thereby creating a cycle in that component. A graph that contains a cycle (*i.e.*, which is not a tree) has a spectral radius larger than or equal to two. This component will, according to Lemma 6.2.3, contribute to an overall spectral radius that is larger than that of a path, which is smaller than 2. □

To prove Theorem 6.2.1, we will use the NP-complete Hamiltonian path problem [133].

Problem 2 (Hamiltonian path problem). *Given a graph $G(N, \mathcal{L})$ with N nodes and L links, a Hamiltonian path is a path that visits every node exactly once. The Hamiltonian path problem is to determine if G contains a Hamiltonian path.*

We are now ready to prove Theorem 6.2.1:

Proof: In our proof we will demonstrate that if we could solve the LSRM problem in polynomial time, then we would also be able to settle the NP-complete Hamiltonian path problem. Assume we have a graph G of $L = N - 1 + m$ links. Removing m links will result in a graph G_m of $N - 1$ links. According to Lemma 6.2.4, a path is the only graph structure of $N - 1$ links that has the smallest largest adjacency eigenvalue and that eigenvalue equals $\lambda_1 = 2 \cos\left(\frac{\pi}{N+1}\right)$. Moreover, a path of $N - 1$ links in a graph of N nodes, is a Hamiltonian path. If, after solving the LSRM problem we obtain $\lambda_1 = 2 \cos\left(\frac{\pi}{N+1}\right)$ (smaller is not possible) then we have found a Hamiltonian path (G_m). If $\lambda_1 > 2 \cos\left(\frac{\pi}{N+1}\right)$, then the original graph G does not contain a Hamiltonian path. The LSRM problem is therefore at least as hard as the Hamiltonian path problem. □

We have to interpret Theorem 6.2.1 with care. Computing the largest eigenvalue can be done in polynomial time. Consequently, the number of possible combinations $\binom{L}{m}$ of m links that we could check (by computing in polynomial time the largest eigenvalue

of the graph G_m resulting after the removal of that specific set of links) is bounded by $O(L^m)$, which is a polynomial function in L . For instance, if $m = 1$, by checking the spectral radius reduction induced by the removal of each of the L links, we can obtain a solution with a complexity of L times the complexity of computing the largest eigenvalue. However, in that case m is fixed and not part of the input N, L, m as defined in problem 1. In other words, m should have been replaced with a fixed integer number in the problem definition to make it clear that m is not part of the input and that its fixed value holds for all problem instances. In problem 1, m is part of the input and, as in our proof, may for instance depend on the number of nodes and links (it makes sense to remove more links in larger networks). The previous argument therefore does not apply to problem 1, which is NP-hard as proved in Theorem 6.2.1. In fact, in our proof $m = L - N + 1$ so that the worst-case complexity of checking all possibilities is $O(L^{L-N+1})$, which is now clearly non-polynomial in the input N, L, m . Similar NP-complete problems, in which the input does not only rely on N and L , but also on another metric k , are the Independent Set problem (defined in problem 4 below) and the Disjoint Connecting Paths problem [133], where k mutually node-disjoint paths need to be found between k corresponding source-destination pairs. This problem also can be solved in polynomial time if k is fixed and thus not part of the input [138], while it is NP-complete if k is part of the input. In general, NP-complete problems that can be solved by algorithms, that are exponential only in the size of a fixed parameter while polynomial in the size of the (remaining) input, are called *fixed-parameter tractable*, because those problems can be solved efficiently for small values of the fixed parameter.

As an example to illustrate the NP-completeness of the LSRM problem, Figures 6.1-6.3 show, in a topology of $N = 10$ nodes and $m = 3$ link removals, that the “best single step strategy” is not always optimal in the end. The “best single step strategy” consists of removing the link that lowers $\lambda_1(A) - \lambda_1(A_1) = y_1$ most in the first step. Next, in the second step, the link that lowers $\lambda_1(A_1) - \lambda_1(A_2) = y_2$ most is removed and finally, in the third step, the link that lowers $\lambda_1(A_2) - \lambda_1(A_3) = y_3$ most is removed. The optimal situation depicts the removal of $m = 3$ for which $\lambda_1(A) - \lambda_1(A_3) = y^*$ is maximal. Hence, $y_1 + y_2 + y_3 \leq y^*$.

In addition, 10^6 instances of Erdős-Rényi (ER) random graphs with $N = 10$ and link density $p = \frac{2 \ln N}{N}$ have been generated. In each instance, the “best single step strategy” and the global optimum have been computed. In 63185 (6,3%) instances, there was no overlap in links, in 332262 (33,2%) ER graphs, there was one link in common, in 97944 (9,8%) ER graphs, we found 2 links in common and in the remaining 506609 (50,7%) ER graphs, all 3 links in the “best single step strategy” were the same as in the global optimum. Moreover, Figure 6.4 illustrates that the global optimum is not always unique. The global optimum may not be unique, as it is possible that the removals of different sets of m links will lead to cospectral or even isomorphic smaller graphs, as indicated in Figure 6.4.

The minimum number m of links that need to be removed from G to ensure that λ_1 in G_m is lowered below some given value χ is

$$m \geq L - \frac{\chi^2 + N - 1}{2},$$

which is derived from the bound [14, (3.48) on p. 54], due to Yuan Hong [139],

$$\lambda_1 \leq \sqrt{2L - N + 1}, \tag{6.1}$$

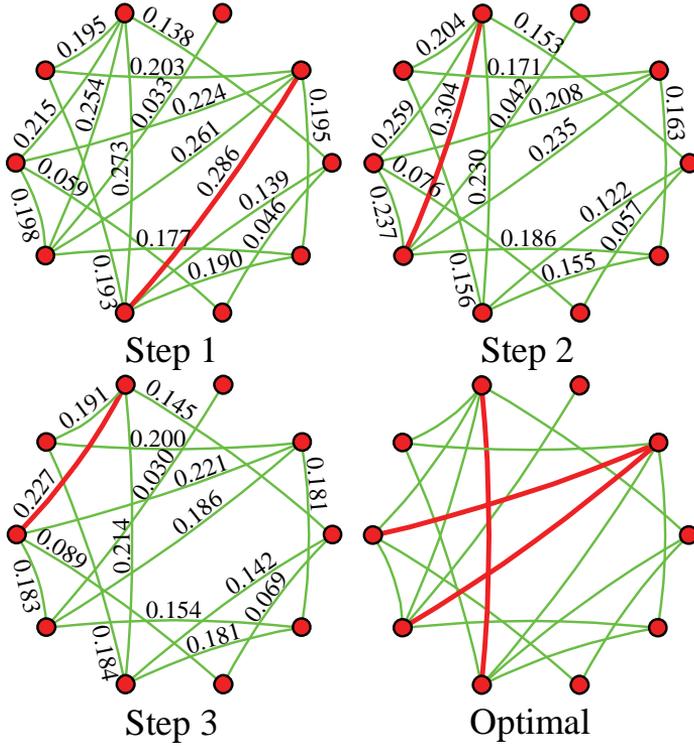


Figure 6.1: An example of a graph with $N = 10$ nodes, where none of the links in the greedy approach appears in the optimal set of links. The numbers indicate the change in largest eigenvalue $\lambda_1(A) - \lambda_1(A \setminus \{l\})$ after removal of link l .

for connected graphs, else $\lambda_1 \leq \sqrt{2L(1 - \frac{1}{N})}$.

6.2.2 The Node Spectral Radius Minimization Problem is NP-hard

Removing nodes to maximally lower the largest eigenvalue may seem an easier problem than removing links. For, when we remove the highest degree node, $\lambda_1(A)$ is likely reduced most (because L is reduced most). This suggestion follows from bounds in [14, p. 48] and the bounds

$$\frac{2L}{N} \sqrt{1 + \frac{\text{Var}[D]}{(E[D])^2}} \leq \lambda_1 \leq \min \left\{ \sqrt{\frac{2L(N-1)}{N}}, d_{\max} \right\}, \tag{6.2}$$

where D is the degree of an arbitrary node in G . Unfortunately, this intuition is wrong. The eigenvalues of the adjacency matrix $A_{l(G)}$ of the line graph $l(G)$ of G and A are related [14, (2.9) on p. 20]. Since links in G are nodes in $l(G)$, and since there is a one-to-one relation between $l(G)$ and G , removing nodes in $l(G)$ according to a certain strategy, results in a corresponding strategy for removing links in G . Since the link spectral radius minimization (LSRM) problem is NP-hard (Theorem 6.2.1), the problem of removing m nodes from a graph G is NP-hard as well. We will provide a proof for

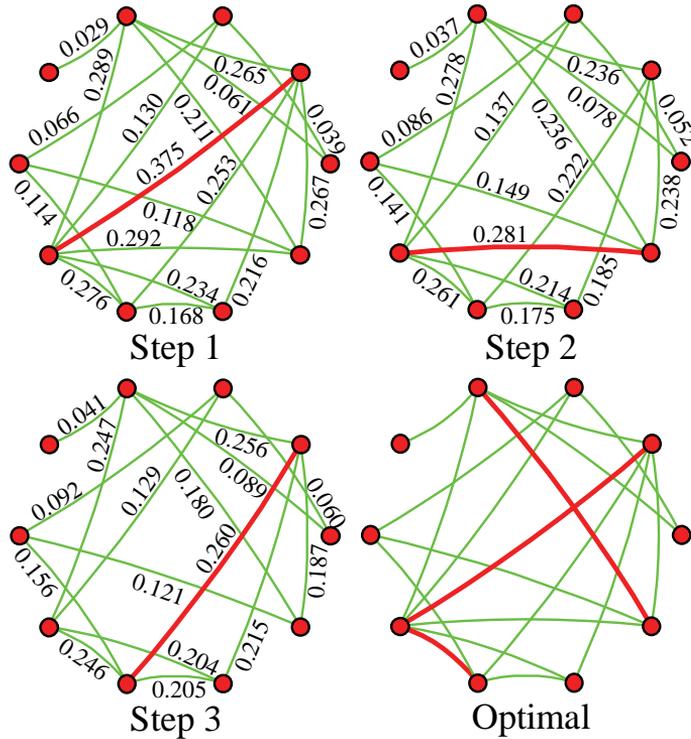


Figure 6.2: An example of a graph with $N = 10$ nodes, where only 1 link in the greedy approach appears in the optimal set of links.

general graphs and subsequently demonstrate that it is also NP-complete in the subclass of line graphs. Let us first formally define the problem:

Problem 3 (Nodal Spectral Radius Minimization (NSRM) problem). *Given a graph $G(\mathcal{N}, \mathcal{L})$ with N nodes and L links, spectral radius $\lambda_1(G)$, and an integer number $m < N$. Which m nodes from the graph G need to be removed, such that the spectral radius of the reduced graph G_m of $N - m$ nodes has the smallest spectral radius out of all possible graphs that can be obtained from G by removing m nodes?*

Theorem 6.2.5. *The NSRM problem is NP-hard.*

We provide a proof by reducing the NP-complete independent set problem [133] to NSRM.

Problem 4 (Independent set problem). *Given a graph $G(\mathcal{N}, \mathcal{L})$ with N nodes and L links and a positive integer $k \leq N$, is there a subset $\mathcal{N}' \subseteq \mathcal{N}$, such that $|\mathcal{N}'| \geq k$ and such that no two nodes in \mathcal{N}' are joined by a link in \mathcal{L} ?*

Proof of Theorem 6.2.5: The lowest spectral radius of a graph equals $\lambda_1(G) = 0$, which is obtained for a graph without any links. Removing nodes that are not part of an independent set, will result in an independent set of nodes that are not linked to each other. Hence, to solve the independent set problem it suffices to remove $m = N - k$ nodes from the graph G , such that the spectral radius of the reduced graph G_m is smallest

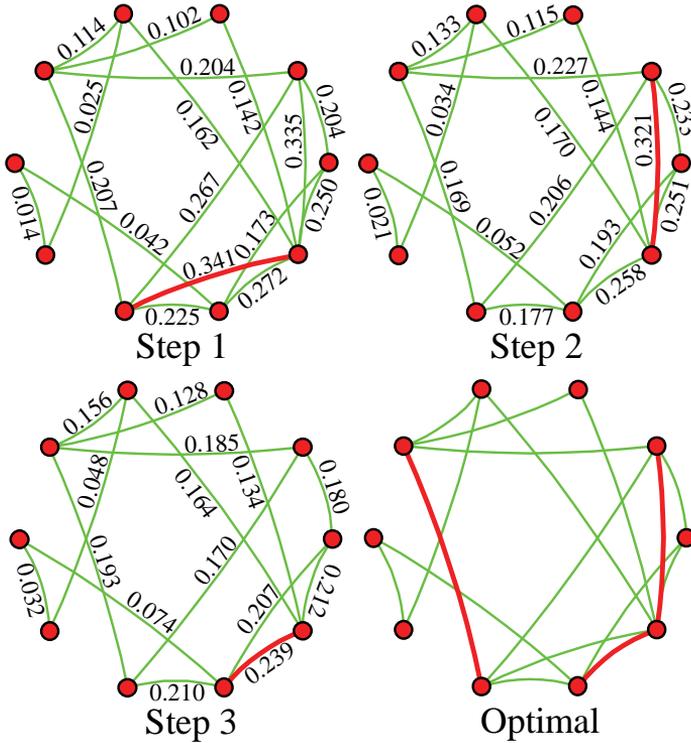


Figure 6.3: An example of a graph with $N = 10$ nodes, where two links in the greedy approach appear in the optimal set of links.

possible. If we get $\lambda_1(G_m) = 0$, then G_m constitutes an independent set of k nodes. If $\lambda_1(G_m) > 0$, then no independent set with at least k nodes exists. \square

Line graphs are a specific class of graphs and not all problems that are NP-complete for general graphs are also NP-complete for line graphs (e.g., according to Roussopoulos [140], the clique problem is not hard in line graphs, while it is an NP-complete problem in general). Hence, we proceed to demonstrate that the NSRM problem remains NP-hard in line graphs. We use similar arguments as for the proof of Theorem 6.2.1. A Hamiltonian path in the graph G corresponds to a path of $N - 1$ nodes in the line graph $l(G)$ of G . A line graph $l(G)$ contains L nodes and can be generated in polynomial time from G . According to Lemma 6.2.2, the graph structure of $N - 1$ nodes that has the smallest largest eigenvalue is the path. Hence, removing $L - N + 1$ nodes from the line graph $l(G)$ such that the spectral radius is reduced most, should correspond to a path of $N - 1$ nodes (if it exists), which corresponds to a Hamiltonian path in G . Solving the NSRM problem in a line graphs $l(G)$ is therefore as hard as finding Hamiltonian paths in the corresponding graph G .

Finally, let l be the removed link that maximizes $\lambda_1(G) - \lambda_1(G \setminus \{l\})$. Let the node n be the transform of link l in the line graph $l(G)$. Then, $\lambda_1(l(G)) - \lambda_1(l(G) \setminus \{n\})$ is not always the maximum. Simulations on 100 Erdős-Rényi random graphs show the “success rate”, the percentage of graphs in which the best link l in G corresponds to the best node n in $l(G)$ in the table 6.1.

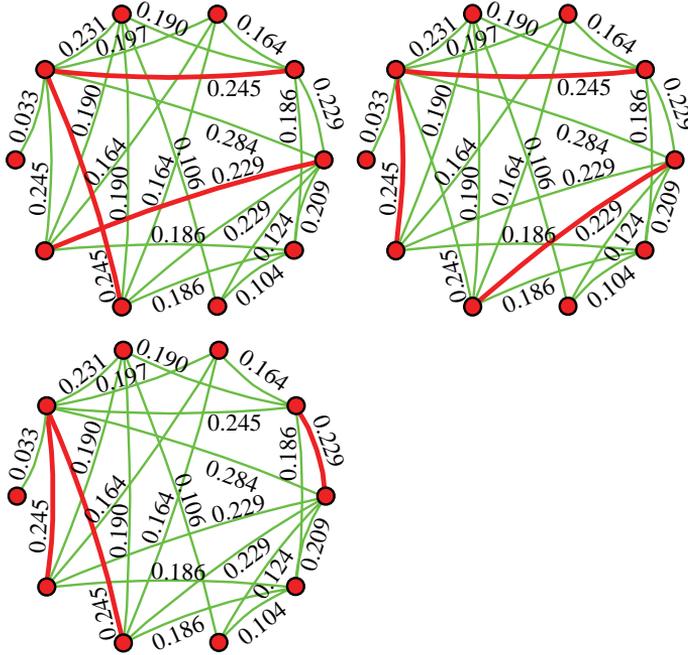


Figure 6.4: An example where the global optimum is not unique. The original $\lambda_1(A) = 5.065310$ and, after removal of three links, the smallest largest eigenvalue is $\lambda_1(A_3) = 4.312414$. Consequently, the largest $\lambda_1(A) - \lambda_1(A_3) = 0.752896$ is obtained after removal of the three red links.

6.3 Removing the Links to Decrease the Spectral Radius

6.3.1 Spectral Graph Theory

We derive a theoretical underpinning to deduce the best heuristic for the LSRM problem. We first introduce the basic spectral graph theory. Let e_j be a base vector in the N -dimensional space, where the i -th component equals $(e_j)_i = \delta_{ij}$ and δ_{ij} is the Kronecker delta, *i.e.* $\delta_{ij} = 1$ if $i = j$ and otherwise, $\delta_{ij} = 0$. Then, the adjacency matrix that represents the single link between nodes i and j equals

$$\hat{A}_{ij} = e_i e_j^T + e_j e_i^T \quad (6.3)$$

	p	0.1	0.2	0.3
N				
10		67%	80%	83%
20		65%	76%	81%
30		59%	76%	81%
40		70%	79%	81%
50		63%	75%	82%
60		67%	82%	86%

Table 6.1: Success rate in line graphs.

Thus, \hat{A}_{ij} equals the zero matrix, except that $(\hat{A}_{ij})_{ij} = (\hat{A}_{ij})_{ji} = 1$. Clearly, $\det(\hat{A}_{ij} - \lambda I) = (-1)^N \lambda^{N-2} (\lambda^2 - 1)$, such that the largest eigenvalue of \hat{A}_{ij} is 1. Also, for any vector z ,

$$z^T \hat{A}_{ij} z = z^T (e_i e_j^T + e_j e_i^T) z = z^T e_i e_j^T z + z^T e_j e_i^T z = 2z_i z_j. \quad (6.4)$$

By invoking $0 \leq (z_i - z_j)^2$, we observe that $2z_i z_j \leq z_i^2 + z_j^2 \leq \sum_{i=1}^N z_i^2 = z^T z$. Hence, when considering normalized vectors such that $z^T z = \|z\|_2^2 = 1$, we obtain the upper bound

$$2z_i z_j \leq 1.$$

After these preliminaries, we now embark on the problem.

The difference $\lambda_1(A) - \lambda_1(A_m)$

With the normalization $x_1^T x_1 = 1$ and $w_1^T w_1 = 1$, the Rayleigh relations [14] become

$$\begin{aligned} \lambda_1(A) &= x_1^T A x_1, \\ \lambda_1(A_m) &= w_1^T A_m w_1. \end{aligned}$$

Writing out the quadratic form

$$\begin{aligned} \lambda_1(A) &= x_1^T A x_1 = \sum_{i=1}^N \sum_{j=1}^N a_{ij}(x_1)_i (x_1)_j = 2 \sum_{i=1}^N \sum_{j=i+1}^N a_{ij}(x_1)_i (x_1)_j \quad (6.5) \\ &= 2 \sum_{l=(i \sim j) \in \mathcal{L}} (x_1)_i (x_1)_j = 2 \sum_{l=1}^L (x_1)_{l^+} (x_1)_{l^-}, \end{aligned}$$

where a link l joins the nodes l^+ and l^- , shows that $\lambda_1(A)$ can be written as a sum of positive products over all links in the graph G .

We now provide a general bound on the difference between the largest eigenvalues in G and $G_m = G \setminus \mathcal{M}_m$, where m links are removed.

Lemma 6.3.1. *For any graph G and $G_m = G \setminus \mathcal{M}_m$, it holds that*

$$2 \sum_{l \in \mathcal{M}_m} (w_1)_{l^+} (w_1)_{l^-} \leq \lambda_1(A) - \lambda_1(A_m) \leq 2 \sum_{l \in \mathcal{M}_m} (x_1)_{l^+} (x_1)_{l^-} \quad (6.6)$$

where x_1 and w_1 are the eigenvectors of A and A_m corresponding to the largest eigenvalues $\lambda_1(A)$ and $\lambda_1(A_m)$, respectively, and where a link l joins the nodes l^+ and l^- .

Proof: Since $A_m = A - \sum_{l \in \mathcal{M}_m} \hat{A}_{l^+ l^-}$ where the left-hand side (or start) of the link l is the node l^+ and the right-hand side (or end) of the link l is the node l^- and with the normalization $x_1^T x_1 = 1$, the Rayleigh relations [14] yield

$$\begin{aligned} \lambda_1(A) &= x_1^T A x_1 = x_1^T \left(A_m + \sum_{l \in \mathcal{M}_m} \hat{A}_{l^+ l^-} \right) x_1 \\ &= x_1^T A_m x_1 + \sum_{l \in \mathcal{M}_m} x_1^T \hat{A}_{l^+ l^-} x_1. \end{aligned}$$

Using (6.4) yields $x_1^T \hat{A}_{l+l-} x_1 = 2(x_1)_{l+} (x_1)_{l-}$ and we arrive at

$$\lambda_1(A) = x_1^T A_m x_1 + 2 \sum_{l \in \mathcal{M}_m} (x_1)_{l+} (x_1)_{l-}.$$

The Rayleigh principle states that, for any normalized vector w with $w^T w = 1$, it holds that $w^T A w \leq \lambda_1(A)$ and equality is only attained if w equals the eigenvector of A belonging to $\lambda_1(A)$. Since x_1 is not the eigenvector of A_m belonging to $\lambda_1(A_m)$, we have that $x_1^T A_m x_1 \leq \lambda_1(A_m)$ and

$$\lambda_1(A) = x_1^T A_m x_1 + 2 \sum_{l \in \mathcal{M}_m} (x_1)_{l+} (x_1)_{l-} \leq \lambda_1(A_m) + 2 \sum_{l \in \mathcal{M}_m} (x_1)_{l+} (x_1)_{l-},$$

from which the upper bound in (6.6) is immediate. When repeating the analysis from the point of view of A_m rather than from A , then

$$\begin{aligned} \lambda_1(A_m) &= w_1^T A_m w_1 = w_1^T \left(A - \sum_{l \in \mathcal{M}_m} \hat{A}_{l+l-} \right) w_1 \\ &= w_1^T A w_1 - 2 \sum_{l \in \mathcal{M}_m} (w_1)_{l+} (w_1)_{l-}. \end{aligned}$$

By invoking the Rayleigh principle again, we arrive at the lower bound. \square

For connected graphs G and G_m , it is known that $\lambda_1(A) - \lambda_1(A_m) > 0$ (see Lemma 7 in [14]). The same conclusion also follows from Lemma 6.3.1 because the Perron-Frobenius theorem [14] states that all vector components of w_1 (and x_1) are positive in a connected graph G_m . Lemma 6.3.1 indicates that, when those m links are removed that maximize $2 \sum_{l \in \mathcal{M}_m} (x_1)_{l+} (x_1)_{l-}$, then the upper bound in (6.6) is maximal, which may lead to the largest possible difference $\lambda_1(A) - \lambda_1(A_m)$. However, those removed links do not necessarily also maximize the lower bound $2 \sum_{l \in \mathcal{M}_m} (w_1)_{l+} (w_1)_{l-}$. Hence, the greedy strategy of removing consecutively the link l with the highest product $(x_1)_{l+} (x_1)_{l-}$ is not necessarily guaranteed to lead to the overall optimum. The fact that the SRM problem is NP-hard, as proved in Section 6.2, underlines this remark.

Lemma 8 in [14] states that

$$\lambda_1(A) - \lambda_1(A_m) \leq \lambda_1(A - A_m).$$

Since $A - A_m = \sum_{l \in \mathcal{M}_m} \hat{A}_{l+l-}$, it remains to find a close upper bound for $\lambda_1\left(\sum_{l \in \mathcal{M}_m} \hat{A}_{l+l-}\right)$. Using the bounds [14, (3.48) on p. 54], gives

$$\begin{aligned} \lambda_1\left(\sum_{l \in \mathcal{M}_m} \hat{A}_{l+l-}\right) &\leq \min\left(\sqrt{2m - N + 1}, d_{\max}(A - A_m)\right) \mathbf{1}_{\{A - A_m \text{ is a connected graph}\}} \\ &\quad + \min\left(\sqrt{2m - \frac{2m}{N}}, d_{\max}(A - A_m)\right) \mathbf{1}_{\{A - A_m \text{ is not a connected graph}\}}. \end{aligned}$$

In general, it is difficult to find sharper bounds (see e.g. [141],[142]). If $m = 2$, then $\lambda_1\left(\sum_{l \in \mathcal{M}_2} \hat{A}_{l+l-}\right) = \sqrt{2}$ when the two links are connected and $\lambda_1\left(\sum_{l \in \mathcal{M}_2} \hat{A}_{l+l-}\right) = 1$ when the two links are disconnected. If $m = 1$, then $\lambda_1\left(\hat{A}_{l+l-}\right) = 1$ and we obtain

$$\lambda_1(A) - \lambda_1(A_1) \leq 1. \tag{6.7}$$

Lemma 6.3.2. *For $m = 1$ link removed from G , equality in (6.7) is only attained for the graph consisting of the complete graph K_N with $N = 2$ nodes and a set of disjoint nodes.*

Proof: Equality in (6.7) combined with (6.6) in Lemma 6.3.1 implies that

$$1 = \lambda_1(A) - \lambda_1(A_1) \leq 2(x_1)_{l^+}(x_1)_{l^-}.$$

Next, from $2(x_1)_{l^+}(x_1)_{l^-} \leq (x_1)_{l^+}^2 + (x_1)_{l^-}^2 \leq x_1^T x_1 = 1$, we conclude that the equality in (6.7) holds if and only if $(x_1)_{l^+} = (x_1)_{l^-} = 1/\sqrt{2}$. Since in such case $(x_1)_{l^+}^2 + (x_1)_{l^-}^2 = 1$, we conclude that all other components of the eigenvector x_1 are equal to zero. Recall that x_1 is the principal eigenvector which, according to the Perron-Frobenius Theory, is positive if G is a connected graph. If G has more than two nodes ($N > 2$), the above argument shows that G must be disconnected with K_2 being the unique component with the largest spectral radius. Therefore, the remaining components must be isolated nodes. \square

Closed walks in subgraphs

Let G be a connected graph with adjacency matrix A . From the decomposition [14, art. 156 on p. 226]

$$A = \sum_{i=1}^n \lambda_i x_i x_i^T,$$

using $x_i^T x_j = 0$ for $i \neq j$ and $x_i^T x_i = 1$ for any i , we have that

$$A^k = \sum_{i=1}^n \lambda_i^k x_i x_i^T.$$

When $k \rightarrow \infty$, the most important term in the sum above is $\lambda_1^k x_1 x_1^T$, provided that G is nonbipartite². In such case, we have $\lambda_1 > |\lambda_i|$ for $i = 2, \dots, n$, and so, for any two nodes u, v of G ,

$$\lim_{k \rightarrow \infty} \frac{(A^k)_{uv}}{\lambda_1^k (x_1)_u (x_1)_v} = \lim_{k \rightarrow \infty} \frac{\sum_{i=1}^n \lambda_i^k (x_i)_u (x_i)_v}{\lambda_1^k (x_1)_u (x_1)_v} = 1 + \sum_{i=2}^n \frac{(x_i)_u (x_i)_v}{(x_1)_u (x_1)_v} \left(\frac{\lambda_i}{\lambda_1} \right)^k = 1.$$

In view of the above, we will deliberately resort to the following *approximation*:

$$\text{For large } k : (A^k)_{uv} \approx \lambda_1^k (x_1)_u (x_1)_v.$$

²In case G is bipartite, let (U, V) be the bipartition of nodes of G . Then $\lambda_n = -\lambda_1$, $(x_n)_u = (x_1)_u$ for $u \in U$ and $(x_n)_v = -(x_1)_v$ for $v \in V$. Both λ_1 and λ_n are simple eigenvalues, so that $\lambda_1 > |\lambda_i|$ for $i = 2, \dots, n-1$. Similarly as above we get

$$\lim_{k \rightarrow \infty} \frac{(A^k)_{u,v}}{\lambda_1^k (x_1)_u (x_1)_v} = 1 + \lim_{k \rightarrow \infty} (-1)^k \frac{(x_n)_u (x_n)_v}{(x_1)_u (x_1)_v}.$$

Obviously, the limit above exists if we restrict k to range over odd or even numbers only, in which case the limit is either 0 or 2, depending on whether u and v belong to the same or different parts of the bipartition. This suggests that the same strategy will extend to bipartite graphs as well, except that the argument will have to take into account the nonexistence of odd closed walks.

Under such approximation, the total number of closed walks of large length k in G is

$$\sum_{u \in V(G)} (A^k)_{uu} \simeq \sum_{u \in V(G)} \lambda_1^k (x_1)_u (x_1)_u = \lambda_1^k \sum_{u \in V(G)} (x_1)_u^2 = \lambda_1^k.$$

We will demonstrate that removing the node u or the link $u \sim v$ with highest vector component $(x_1)_u$ or highest vector component product $(x_1)_u (x_1)_v$ will decrease $\lambda_1(A)$ most.

The link whose deletion reducing λ_1 most

We want to find out the deletion of which link $u \sim v$ mostly reduces the number of closed walks in G of some large length k ?

For fixed u, v and k , let W_t denote the number of closed walks of length k which start at some node w and contain the link $u \sim v$ at least t times, $t \geq 1$. Suppose that in such walk, the link $u \sim v$ appears at positions $1 \leq l_1 \leq l_2 \leq \dots \leq l_t \leq k$ in the sequence of links on the walk, and let $u_{i,0}$ and $u_{i,1}$ be the first and the second node of the i th appearance of uv in the walk. Obviously, either $(u_{i,0}, u_{i,1}) = (u, v)$ or $(u_{i,0}, u_{i,1}) = (v, u)$. Then

$$\begin{aligned} W_t &= \sum_{w \in V} \sum_{l_1 \leq \dots \leq l_t} (A^{l_1-1})_{wu_{1,0}} \left(\prod_{i=2}^t (A^{l_i-l_{i-1}-1})_{u_{i-1,1}u_{i,0}} \right) (A^{k-l_t-1})_{u_{t,1}w} \\ &\simeq \sum_{w \in V} \sum_{l_1 \leq \dots \leq l_t} \lambda_1^{l_1-1} (x_1)_w (x_1)_{u_{1,0}} \\ &\quad \times \left(\prod_{i=2}^t \lambda_1^{l_i-l_{i-1}-1} (x_1)_{u_{i-1,1}} (x_1)_{u_{i,0}} \right) \lambda_1^{k-l_t-1} (x_1)_{u_{t,1}} (x_1)_w \\ &= \sum_{w \in V} (x_1)_w^2 \sum_{l_1 \leq \dots \leq l_t} \lambda_1^{k-t} \prod_{i=1}^t ((x_1)_{u_{i,0}} (x_1)_{u_{i,1}})^2 \\ &= \binom{k}{t} \lambda_1^{k-t} (2(x_1)_u (x_1)_v)^t. \end{aligned}$$

The term $2(x_1)_u (x_1)_v$ appears in the last equation because there are two ways to choose $(x_{u_{i,0}}, x_{u_{i,1}})$ for each $i = 1, \dots, t$.

Now, the number of walks affected by deleting the link $u \sim v$ is equal to

$$\begin{aligned} W^{uv} &= \sum_{t \geq 1} (-1)^{t-1} W_t \\ &= \sum_{t \geq 1} (-1)^{t-1} \binom{k}{t} \lambda_1^{k-t} (2(x_1)_u (x_1)_v)^t \\ &= \lambda_1^k - \sum_{t \geq 0} (-1)^t \binom{k}{t} \lambda_1^{k-t} (2(x_1)_u (x_1)_v)^t \\ &= \lambda_1^k - (\lambda_1 - 2(x_1)_u (x_1)_v)^k. \end{aligned}$$

The last function is increasing in $(x_1)_u (x_1)_v$ in the interval $[0, \lambda_1/2]$, and so most closed walks are destroyed when we remove the link with the largest product of principal eigenvector components. Thus, the spectral radius is decreased the most in such case as well.

Assortativity and lower bounds for λ_1

A lower bound of the largest adjacency eigenvalue $\lambda_1 \geq \frac{N_3}{N_2}$ has been proved in [40], where N_k is the total number of walks of length k . The lower bound $\frac{N_3}{N_2}$ appeared earlier as an approximation in [38] of the largest adjacency eigenvalue λ_1 , and it is a perfect linear function of assortativity ρ_D [40].

Let us first look at the decrease of $\frac{N_3}{N_2}$ by a link removal. We know [40] that

$$\frac{N_3}{N_2} = \frac{\sum_{i=1}^N d_i^3 - \sum_{i \sim j} (d_i - d_j)^2}{\sum_{i=1}^N d_i^2}.$$

We denote N_3 and N'_3 as the number of 3 hop walks in the original graph G and in the graph $G \setminus \{l_{ij}\}$ with one link $l = i \sim j$ less, respectively. Then, we have that

$$\begin{aligned} \Delta_3 &= N_3 - N'_3 \\ &= d_i^3 + d_j^3 - (d_i - 1)^3 - (d_j - 1)^3 - (d_i - d_j)^2 - \sum_{l \in \mathcal{N}(i), l \neq j} (d_l - d_i)^2 - (d_l - d_i + 1)^2 \\ &\quad - \sum_{l \in \mathcal{N}(j), l \neq i} (d_l - d_j)^2 - (d_l - d_j + 1)^2, \end{aligned}$$

where d_i is the degree of node i in the original graph G and $\mathcal{N}(i)$ is the set of the neighbors of node i . The decrease Δ_3 can be simplified as

$$\begin{aligned} \Delta_3 &= 2 - 3(d_i + d_j) + 3(d_i^2 + d_j^2) - (d_i - d_j)^2 + \sum_{l \in \mathcal{N}(i), k \neq j} (2d_l - 2d_i + 1) \\ &\quad + \sum_{l \in \mathcal{N}(j), l \neq i} (2d_l - 2d_j + 1) \\ &= 2(d_i^2 + d_j^2) + 2d_i d_j + 2 - 3(d_i + d_j) + (d_i + d_j - 2) - 2d_i(d_i - 1) \\ &\quad - 2d_j(d_j - 1) + \sum_{l \in \mathcal{N}(i), k \neq j} 2d_l + \sum_{l \in \mathcal{N}(j), k \neq i} 2d_l \\ &= 2d_i d_j + \sum_{l \in \mathcal{N}(i), k \neq j} 2d_l + \sum_{l \in \mathcal{N}(j), k \neq i} 2d_l \\ &= 2d_i d_j + 2(s_i + s_j) - 2(d_i + d_j), \end{aligned}$$

where

$$s_i = \sum_{l \in \mathcal{N}(i)} d_l \tag{6.8}$$

is the total degree of all the direct neighbors of a node i . Similarly, the decrease in the number of two hop walks is denoted as

$$\Delta_2 = N_2 - N'_2 = 2(d_i + d_j - 1).$$

Note that Δ_2 and Δ_3 are only functions of a local property, *i.e.* the degree d_i and d_j of the two end nodes of a link l_{ij} . The complexity of computing Δ_3 or Δ_2 for all linked node pairs is $O(N^2)$ in a dense graph, which is the worst case.

6.3.2 Strategies to Minimize the Largest Eigenvalue by Link Removal

This section discusses and compares various strategies in Figure 6.5, denoted by S .

The first strategy, as suggested in Section 6.3.1, is to remove the link with maximum product of the eigenvector components. Specifically, this strategy is denoted by $S = x_i x_j$ instead of $S = (x_1)_i (x_1)_j$ to simplify the notation in the figures, and it removes that link $l = i \sim j$ for which $(x_1)_i (x_1)_j$ is maximal.

Section 6.3.1 hints that the spectral radius is possibly decreased the most by a link removal that either reduces $S = \frac{N_3}{N_2}$ or the assortativity $S = \rho_D$ the most. Strategy $S = \frac{N_3}{N_2}$ will remove the link such that $\frac{N_3 - \Delta_3}{N_2 - \Delta_2}$ is minimized.

The other considered strategies $S = d_i d_j$ and $S = d_i + d_j$ remove that link $l = i \sim j$ with largest sum or product of the degrees of the link's end points, whereas the strategies $S = s_i + s_j$ and $S = s_i s_j$ remove the link with the largest sum or product of the total degree s_i of the neighbors at both end points. Finally, we also considered the strategy $S = \text{betweenness}$, that removes the link with highest link betweenness, *i.e.* the number of shortest paths between all node pairs that traverse the link.

We define the performance measure Ξ_S of a particular link removal strategy S by

$$\Xi_S = (\lambda_1(A) - \lambda_1(A_1))_{\text{optimal}} - (\lambda_1(A) - \lambda_1(A_1))_{\text{Strategy } S}.$$

Figure 6.5 compares the above explained strategies. Figure 6.5 confirms that strategy $S = x_i x_j$ is superior to all other strategies. There is a very small difference between the strategies $S = d_i + d_j$ and $S = d_i d_j$ and between $S = s_i + s_j$ and the corresponding product $S = s_i s_j$. In both cases the product strategy is slightly better (but the difference is not observable from Figure 6.5).

Another strategy is to remove the link that possibly disconnects the graph G into two disjoint graphs G_1 and G_2 . However, this strategy is not always optimal as illustrated in Figure 6.6. Only when both G_1 and G_2 are the same, we found that the removal of the connecting link induces the largest decrease in $\Delta\lambda_1$. Since this strategy cannot be applied always, we have further ignored this strategy.

Removing $m > 1$ links

In this section, we investigate the behavior of the several strategies when more than one link is removed. We generated 10^4 Erdős-Rényi graphs with $N = 10$ nodes and $L = 20$ links, of which about two percent are disconnected. From each of the generated graphs, all the links are removed one by one following the different “greedy” strategies. We compare the decrease in λ_1 for each strategy to the optimal solution found by removing all possible combinations of m links. In Figure 6.7, the percentage of agreement between the greedy strategies and the optimal strategy is shown.

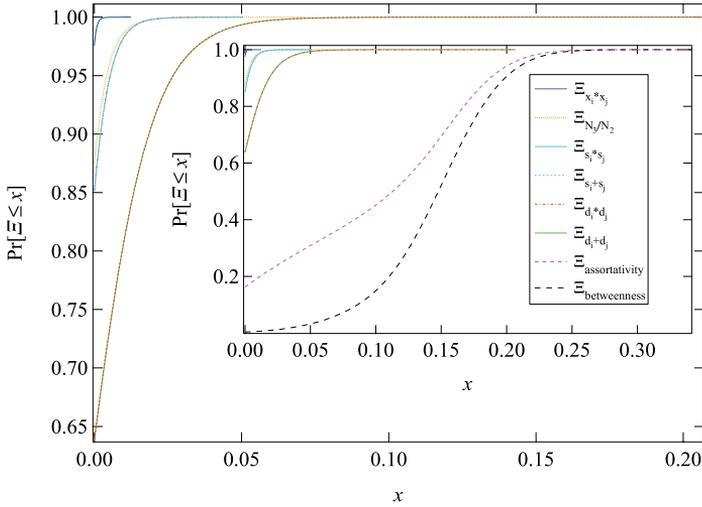


Figure 6.5: Various strategies applied to 10^6 instances of ER graphs with $N = 20$ and $p = 2 \ln N/N$. The insert shows two additional strategies “assortativity” and “betweenness” that are clearly worse than the others.

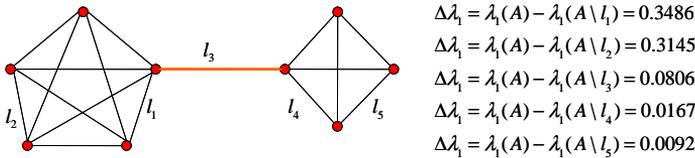


Figure 6.6: All possible $\Delta\lambda_1$ are computed when one link is removed.

Figure 6.7 illustrates that strategy $S = \max_{1 \leq (i,j) \leq N} (x_1)_i (x_1)_j$ is nearly always (except for $m = 13$) superior to strategy $S = N_3/N_2$ and $S = s_i s_j$, which agrees with the theory in Section 6.3.1. Figure 6.7 exhibits a regime change from $m = 10$ on, where the connectivity of the graphs starts to decrease rapidly.

The peculiar regime for $m > 10$ can be understood as follows. The optimal solution for $m = 10$ removals is a circuit, if the original graph contains a single connected circuit on $N = 10$ nodes. If strategy $S = \max_{1 \leq (i,j) \leq N} (x_1)_i (x_1)_j$ finds the optimal solution for $m = 10$ removals, the only possible solution for $m = 11$ removals is to cut the circuit to form a path. This is also the optimal solution. The eigenvector components of a path graph are symmetrical around the node(s) in the middle of the path and are maximal for the center node(s). Strategy $S = \max_{1 \leq (i,j) \leq N} (x_1)_i (x_1)_j$ will, for the next link removal, cut the path in the middle. The resulting graph is also the optimal solution. In the next step, however, the strategy will cut one of the paths in two, resulting in three paths of lengths one, two and four links, respectively. The optimal solution for $m = 13$ link removals consists of a graph with three paths of lengths two and one of length three. This graph can never be formed by strategy $S = \max_{1 \leq (i,j) \leq N} (x_1)_i (x_1)_j$ starting from a circuit. The optimal solution for $m = 14$ consists of two paths of length two and two paths of length one, which can be obtained in many different ways, including cutting the longest path of the solution for $m = 13$. In almost 98% of the cases this solution is found by strategy $S = \max_{1 \leq (i,j) \leq N} (x_1)_i (x_1)_j$. The high success rate means, at the same time,

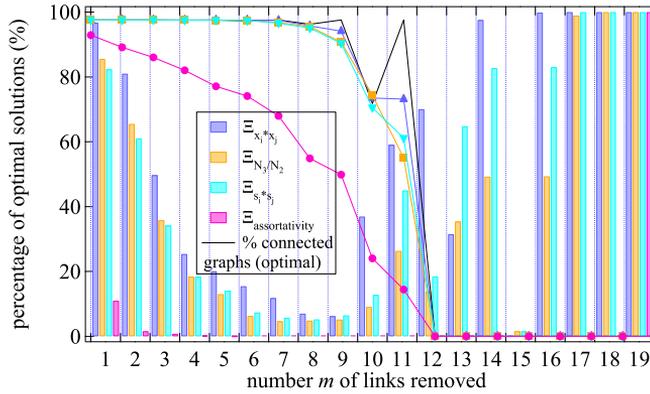


Figure 6.7: Four strategies compared with the global optimum as a function of the number m of removed links in ER random graphs with $N = 10$ nodes and $L = 20$ links, where 10^4 instances are generated. The lines show the percentage of connected graphs per strategy after the removal of m links.

that the optimal solution for $m = 15$ is almost never found because it cannot be reached from the optimal solution of $m = 14$ by another link removal, regardless of the followed strategy. The weaker performance of strategy $S = s_i s_j$ for $m = 12$ can be explained by considering the optimal solution for $m = 11$ which is a path of nine links. Strategy $S = s_i s_j$ removes the link that has the maximum product of the one hop neighbors of its endpoints. Since a path has an even degree distribution, except for the endpoints, the five links that form the center of the path have an equal probability of being removed. Consequently, the optimal solution for $m = 11$ will result in the optimal solution for $m = 12$ only one in five times. The other four possibilities lead to a graph with either a combination of a path of length two and a path of length six or a combination of a path of length three and a path of length five. Both these graphs will give the optimal solution for $m = 13$ link removals, which explains the increased success rate for $m = 13$.

At $m = 15$, the graph consists of 5 links and $N = 10$ nodes, configured in separated “cliques” K_2 (*i.e.* line segments) and the largest eigenvalue is minimal at $\lambda_1 = 1$. For $m > 15$, the strategies are all the same: a clique K_2 (*i.e.* disjoint link) is removed.

Figure 6.8 illustrates four strategies on a typical instance of a network with $N = 10$ and $L = 20$ links. While the strategy $S = \text{assortativity}$ clear underperforms, the three other strategies $S = x_i x_j$, $S = N_3/N_2$ and $S = s_i s_j$ are competitive: for small m , the strategy $S = x_i x_j$ excels (as shown in Figure 6.7), but for larger m the others can outperform. Again, this phenomenon is characteristic for an NP-complete problem, where the whole previous history of links removals affects the current link removal. The considered strategies (except for the global optimum one) are greedy and only optimize the current link removal, irrespective of the way in which the current graph G_m is obtained previously.

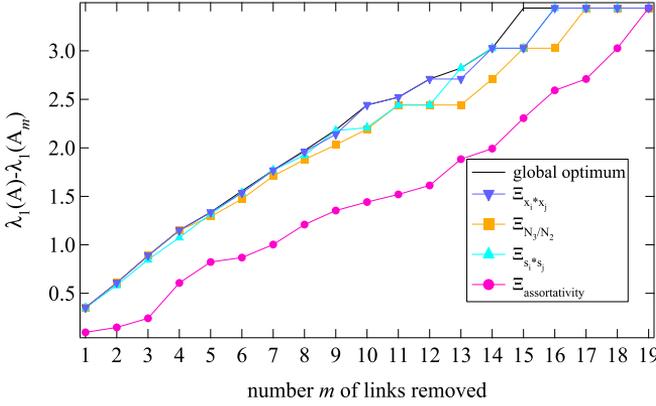


Figure 6.8: The performance $\lambda_1(A) - \lambda_1(A_m)$ of four strategies versus m link removals in a typical instance of a graph with $N = 10$ and $L = 20$ links.

6.3.3 Scaling Law of $(\lambda_1(A) - \lambda_1(A_m))_{\text{optimal}}$

Another observation from Figure 6.8 is that

$$\Delta\lambda_m|_{\text{optimal}} = \lambda_1(A) - \lambda_1(A_m)|_{\text{optimal}} = O(m^\beta), \tag{6.9}$$

where $\beta \leq 1$. In other words, we conjecture that the scaling of $\lambda_1(A) - \lambda_1(A_m)$ with m is sublinear in m (for non-regular graphs) and that the coefficient β is likely a function of the type of graph. Obviously, $\Delta\lambda_m = 0$, when $m = 0$. Applying the upper bound (6.1) to $\lambda_1(A_m)$ shows that

$$\begin{aligned} \Delta\lambda_m &\geq \lambda_1(A) - \sqrt{1 - \frac{1}{N}\sqrt{2L - 2m}} \\ &\geq \lambda_1(A) - \sqrt{1 - \frac{1}{N}\sqrt{2L}} + \sqrt{1 - \frac{1}{N}\sqrt{2m}} = O(m^{1/2}). \end{aligned}$$

On the other hand, if G_m is a regular graph, then

$$\Delta\lambda_m = \lambda_1(A) - \frac{2L - 2m}{N} = O(m).$$

In particular, if G and G_m are regular graphs, then

$$\Delta\lambda_m = \frac{2m}{N}. \tag{6.10}$$

These arguments illustrate that $\frac{1}{2} < \beta \leq 1$. Figure 6.8 shows that $\lambda_1(A) - \lambda_1(A_m)|_{\text{optimal}}$ is likely close to $\beta = 1$, which suggests that the optimal way to remove m links is to make G_m as regular as possible, because the lowest possible $\lambda_1(A_m)$ with given N and $L - m$ is obtained for a regular graph (as follows from the Rayleigh inequality $\lambda_1(A) \geq \frac{2L}{N}$).

While the law (6.9) is difficult to prove in general, we provide evidence by computing the decrease in λ_1 when m random links are removed in the class of Erdős-Rényi random

graphs $G_p(N)$. For sufficiently large Erdős-Rényi random graphs $G_p(N)$, we know [14] that

$$E[\lambda_1] = (N-2)p + 1 + O\left(\frac{1}{\sqrt{N}}\right).$$

When m random links are removed from $G_p(N)$, we again obtain an Erdős-Rényi random graph with link density

$$p^* = \frac{L-m}{\binom{N}{2}}.$$

Hence,

$$\begin{aligned} E[\Delta\lambda_m] &= E[\lambda_1(G_p(N))] - E[\lambda_1(G_{p^*}(N))] \\ &= (N-2)(p-p^*) + R_p(N), \end{aligned}$$

where the error term $R_p(N)$ is unknown. Assuming that $R_p(N)$ is negligibly small, we find, for sufficiently high N ,

$$E[\Delta\lambda_m] \simeq \frac{(N-2)m}{\binom{N}{2}} = \frac{2m}{N} - \frac{2m}{N(N-1)}.$$

Thus, the average decrease in $\lambda_1(A) - \lambda_1(A_m)$ after removing m random links in $G_p(N)$ is approximately, for large N ,

$$E[\Delta\lambda_m] \simeq \frac{2m}{N}, \quad (6.11)$$

which is close to (6.10) for regular graphs.

For $m=1$, simulations on various types of graphs in Figures 6.9 and 6.10 suggest the scaling law

$$(\lambda_1(A) - \lambda_1(A_1))_{\text{optimal}} = \frac{\alpha}{N}, \quad (6.12)$$

where α is graph specific. In other words, $N\Delta\lambda_1 = \alpha$ is independent of the size of the graph.

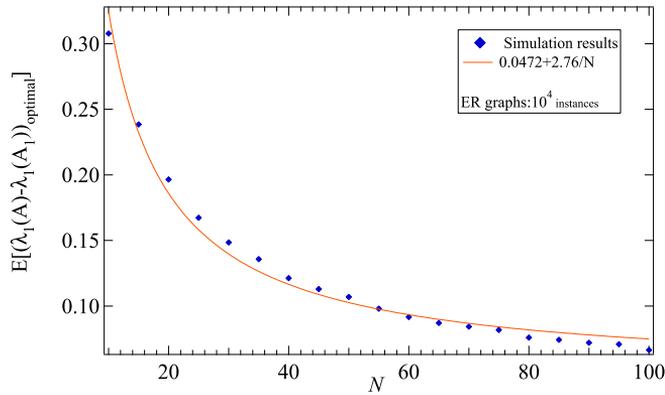


Figure 6.9: The scaling law of $(\lambda_1(A) - \lambda_1(A_1))_{\text{optimal}}$ for ER random graphs as a function of N .

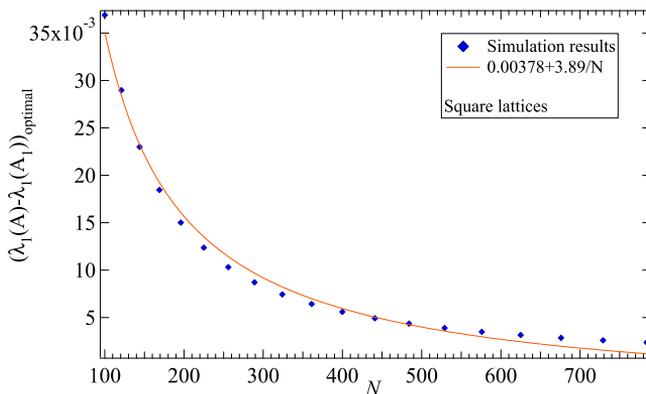


Figure 6.10: The scaling law of $(\lambda_1(A) - \lambda_1(A_1))_{optimal}$ for square lattices as a function of N .

Ignoring the asymptotic nature of the analysis that led to (6.11), we observe that, for $m = 1$, a maximum occurs at $N = 2$. Figure 6.11 shows the pdf of $\Delta\lambda$ for Erdős-Rényi random graphs, where for each curve 10^6 ER graphs have been created in which one random link was removed. The simulations agree with $E[\Delta\lambda] \simeq \frac{2}{N}$ and indicate that $\text{Var}[\Delta\lambda] \simeq 13 + 2E[\Delta\lambda]$. Since a random link removal is inferior to the removal of the optimal link, Figure 6.9 indeed illustrates that the coefficient of the inverse N scaling law $\alpha_{G_p(N)} \simeq 2.75 > 2$. Figure 6.10 shows that $\alpha_{lattice} \simeq 3.9 > \alpha_{G_p(N)} \simeq 2.75$, which may indicate that deviations from regularity causes λ_1 to decrease more.

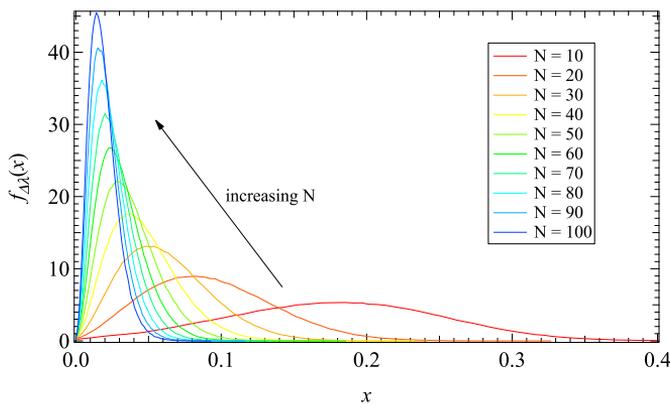


Figure 6.11: The probability density function of $\Delta\lambda_1$ for ER random graphs of several sizes N .

6.4 Removing the Nodes to Decrease the Spectral Radius

6.4.1 The Node Whose Deletion Reducing λ_1 Most

In order to find the node whose deletion reduces λ_1 most, we will consider the *equivalent* question: which deleted node u reduces the number of closed walks in G for some large

length k most?

The closed walks in sub-graph have been discussed in Section 6.3.1. Of course, the number of closed walks of length k which start at node u is equal to $(A^k)_{uu} \approx \lambda_1^k (x_1)_u^2$. When we delete node u from G , then, besides the closed walks which start at u , we also destroy the closed walks which start at another node v , but which contain u as well. Any such closed walk that starts at v may contain several occurrences of u .

For fixed u , k and v , let W_t denote the number of closed walks of length k which start at v and which contain u at least t times, $t \geq 1$. Suppose that in such a walk, node u appears after l_1 steps, after $l_1 + l_2$ steps, after $l_1 + l_2 + l_3$ steps, and so on, the last appearance accounted after $l_1 + \dots + l_t$ steps. Here $l_1, \dots, l_t \geq 1$. Moreover, u must appear for the last time after at most $k - 1$ steps (after k steps we are back at v), thus we may also introduce $l_{t+1} = k - (l_1 + \dots + l_t)$ and ask that $l_{t+1} \geq 1$. Then, we have

$$\begin{aligned} W_t &= \sum_{l_1, \dots, l_{t+1}} (A^{l_1})_{vu} (A^{l_2})_{uv} \dots (A^{l_t})_{uv} (A^{l_{t+1}})_{vu} \\ &\simeq \sum_{l_1, \dots, l_{t+1}} \lambda_1^k (x_1)_v^2 (x_1)_u^{2t} = \lambda_1^k (x_1)_v^2 (x_1)_u^{2t} \sum_{\sum_{j=1}^{t+1} l_j = k; l_j \geq 1} 1. \end{aligned}$$

Introducing $l'_1 = l_1 - 1, \dots, l'_{t+1} = l_{t+1} - 1$, the last sum is equal to the number of nonnegative solutions to

$$l'_1 + l'_2 + \dots + l'_t + l'_{t+1} = k - t - 1,$$

which is, in turn, equal to $\binom{(k-1-t)+t}{t} = \binom{k-1}{t}$. Therefore,

$$W_t \simeq \binom{k-1}{t} \lambda_1^k (x_1)_v^2 (x_1)_u^{2t}.$$

Consider now a closed walk of length k starting at v which contains u exactly j times. Such walk is counted j times in W_1 , $\binom{j}{2}$ times in W_2 , $\binom{j}{3}$ times in W_3 , \dots , $\binom{j}{j}$ times in W_j , and using the well-known equality

$$1 = \sum_{t \geq 1} (-1)^{t-1} \binom{j}{t}$$

we see that this closed walk is counted exactly once in the expression

$$W^v = W_1 - W_2 + W_3 - \dots + (-1)^{t-1} W_t + \dots$$

Thus, W^v represents the number of closed walks of length k starting at v which will be affected by deleting u . From the above expression for W_t , we have

$$\begin{aligned} W^v &\simeq \sum_{t \geq 1} (-1)^{t-1} \binom{k-1}{t} \lambda_1^k (x_1)_v^2 (x_1)_u^{2t} \\ &= -\lambda_1^k (x_1)_v^2 \sum_{t \geq 1} \binom{k-1}{t} (-x_1)_u^{2t} \\ &= \lambda_1^k (x_1)_v^2 \left[1 - (1 - (x_1)_u^2)^{k-1} \right]. \end{aligned}$$

Therefore, the total number of closed walks of length k destroyed by deleting u is equal to

$$\begin{aligned}
 W &\simeq \lambda_1^k (x_1)_u^2 + \sum_{v \neq u} W^v \\
 &= \lambda_1^k (x_1)_u^2 + \lambda_1^k \sum_{v \neq u} (x_1)_v^2 \left[1 - (1 - (x_1)_u^2)^{k-1} \right] \\
 &= \lambda_1^k \left[(x_1)_u^2 + (1 - (x_1)_u^2) \left[1 - (1 - (x_1)_u^2)^{k-1} \right] \right] \\
 &= \lambda_1^k \left[1 - (1 - (x_1)_u^2)^k \right].
 \end{aligned}$$

The last function is increasing in $(x_1)_u$ in the interval $[0, 1]$, and so we conclude that most closed walks are destroyed when we remove the node with the largest principal eigenvector component. Hence, the spectral radius (see (6.5)) is decreased the most in such case as well.

6.4.2 The Bounds for λ_1 When Nodes are Removed

Many inequalities for the spectral radius have been published (see e.g. [142] and [14]). The search to improve the bounds for the spectral radius will continue due to the intimate relation with dynamic processes such as epidemics and synchronization in networks as explained in [108]. Our main result here is:

Theorem 6.4.1. *For any graph G and corresponding graph $G_m(\mathcal{N}) = G \setminus \mathcal{N}_m$, obtained from G by removing the set \mathcal{N}_m of m nodes, it holds that*

$$\left(1 - 2 \sum_{n \in \mathcal{N}_m} (x_1)_n^2 \right) \lambda_1(A) + \sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} (x_1)_i (x_1)_j \leq \lambda_1(A_m(\mathcal{N})) \leq \lambda_1(A) \quad (6.13)$$

where x_1 is the eigenvector of A corresponding to the largest eigenvalue $\lambda_1(A)$. In particular, if $m = 1$, then

$$\left(1 - 2(x_1)_n^2 \right) \lambda_1(A) \leq \lambda_1(A_1(\mathcal{N})) \leq \lambda_1(A) \quad (6.14)$$

Proof: After removing a node n from graph G , we obtain $A_1(\mathcal{N})$, which is a $(N - 1) \times (N - 1)$ matrix

$$A_1(\mathcal{N}) = \begin{bmatrix} a_{11} & \cdots & a_{1(n-1)} & a_{1(n+1)} & \cdots & a_{1N} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{(n-1)1} & \cdots & a_{(n-1)(n-1)} & a_{(n-1)(n+1)} & \cdots & a_{(n-1)N} \\ a_{(n+1)1} & \cdots & a_{(n+1)(n-1)} & a_{(n+1)(n+1)} & \cdots & a_{(n+1)N} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{N1} & \cdots & a_{N(n-1)} & a_{N(n+1)} & \cdots & a_{NN} \end{bmatrix}.$$

Consider the $N \times N$ matrix

$$\widetilde{A}_1(\mathcal{N}) = \begin{bmatrix} a_{11} & \cdots & a_{1(n-1)} & 0 & a_{1(n+1)} & \cdots & a_{1N} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{(n-1)1} & \cdots & a_{(n-1)(n-1)} & 0 & a_{(n-1)(n+1)} & \cdots & a_{(n-1)N} \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ a_{(n+1)1} & \cdots & a_{(n+1)(n-1)} & 0 & a_{(n+1)(n+1)} & \cdots & a_{(n+1)N} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{N1} & \cdots & a_{N(n-1)} & 0 & a_{N(n+1)} & \cdots & a_{NN} \end{bmatrix},$$

which has the same largest eigenvalue as $A_1(\mathcal{N})$. In fact, all eigenvalues of $A_1(\mathcal{N})$ are the same as in $\widetilde{A}_1(\mathcal{N})$, that possesses an additional zero eigenvalue. In the following deduction, we likewise consider $\widetilde{A}_1(\mathcal{N})$ instead of $A_1(\mathcal{N})$ in order to have the dimension equal to $N \times N$. The principal eigenvector w_1 corresponding to $\lambda_1(A_m(\mathcal{N}))$ is also extended to a vector with N components, where the components corresponding to the removed nodes are all zeros.

The Rayleigh principle states that $x^T A x \leq \lambda_1(A)$ for any normalized vector x with $x^T x = 1$ and equality is only attained when $x = x_1$. Since x_1 is an eigenvector of A , but not necessarily an eigenvector of $\widetilde{A}_1(\mathcal{N})$ belonging to $\lambda_1(\widetilde{A}_1(\mathcal{N}))$, we have that $\lambda_1(\widetilde{A}_1(\mathcal{N})) \geq x_1^T (\widetilde{A}_1(\mathcal{N})) x_1$, where

$$x_1^T (\widetilde{A}_1(\mathcal{N})) x_1 = x_1^T A x_1 - x_1^T (A - \widetilde{A}_1(\mathcal{N})) x_1 = \lambda_1(A) - x_1^T (A - \widetilde{A}_1(\mathcal{N})) x_1. \quad (6.15)$$

It remains to compute $x_1^T (A - \widetilde{A}_1(\mathcal{N})) x_1$. We can write

$$A - \widetilde{A}_1(\mathcal{N}) = a_n \cdot e_n^T + e_n \cdot a_n^T,$$

where a_n is the column vector $(a_{n1}, a_{n2}, \dots, a_{nN})^T$ and e_n is the n -th basis column vector $(0, 0, \dots, 1, \dots, 0)^T$, where only the n -th component is 1. Hence,

$$\begin{aligned} x_1^T (A - \widetilde{A}_1(\mathcal{N})) x_1 &= x_1^T (a_n \cdot e_n^T + e_n \cdot a_n^T) x_1 \\ &= x_1^T a_n e_n^T x_1 + x_1^T e_n a_n^T x_1 = 2(x_1)_n \sum_{i=1}^N (x_1)_i a_{in}. \end{aligned}$$

The eigenvalue equation written for the component n yields

$$\sum_{i=1}^N (x_1)_i a_{in} = \lambda_1(A) (x_1)_n,$$

so that we arrive at

$$x_1^T (A - \widetilde{A}_1(\mathcal{N})) x_1 = 2(x_1)_n^2 \lambda_1(A). \quad (6.16)$$

Introduced in (6.15) yields the lower bound in (6.14).

We repeat the analysis from the point of view of $\widetilde{A}_1(\mathcal{N})$. Since w_1 is an eigenvector of $\widetilde{A}_1(\mathcal{N})$, but not necessarily an eigenvector of A belonging to $\lambda_1(A)$, we have $\lambda_1(A) \geq$

$w_1^T A w_1$. Similarly as above,

$$\begin{aligned}\lambda_1(A) &\geq w_1^T \widetilde{A}_1(\mathcal{N}) w_1 + w_1^T \left(A - \widetilde{A}_1(\mathcal{N}) \right) w_1 \\ &= \lambda_1(\widetilde{A}_1(\mathcal{N})) + w_1^T \left(A - \widetilde{A}_1(\mathcal{N}) \right) w_1 \\ &= \lambda_1(\widetilde{A}_1(\mathcal{N})) + 2\lambda_1(\widetilde{A}_1(\mathcal{N}))(w_1)_n^2,\end{aligned}\tag{6.17}$$

from which, with $\sum_{i=1}^N (w_1)_i a_{in} = \lambda_1(\widetilde{A}_1(\mathcal{N}))(w_1)_n$ and $a_n = 0$ in $\widetilde{A}_1(\mathcal{N})$ so that $(w_1)_n = 0$, the upper bound in (6.14) follows.

Next, we extend inequality (6.15) in case m nodes are removed,

$$x_1^T (A - A_m(\mathcal{N})) x_1 = x_1^T \left(\sum_{n \in \mathcal{N}_m} a_n \cdot e_n^T + \sum_{n \in \mathcal{N}_m} e_n \cdot a_n^T - \sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} e_i e_j^T \right) x_1,$$

and obtain

$$\begin{aligned}\lambda_1(A_m(\mathcal{N})) &\geq \lambda_1(A) - x_1^T (A - A_m(\mathcal{N})) x_1 \\ &= \lambda_1(A) - 2\lambda_1(A) \sum_{n \in \mathcal{N}_m} (x_1)_n^2 + \sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} (x_1)_i (x_1)_j.\end{aligned}\tag{6.18}$$

Similarly, when repeating the analysis from the point of view of $A_m(\mathcal{N})$ rather than from A , we can also extend inequality (6.17) in case m nodes are removed. With $\lambda_1(A) \geq w_1^T (A) w_1$, we achieve

$$\begin{aligned}\lambda_1(A) &\geq \lambda_1(A_m(\mathcal{N})) - w_1^T (A_m(\mathcal{N}) - A) w_1 \\ &= \lambda_1(A_m(\mathcal{N})) + 2\lambda_1(A_m(\mathcal{N})) \sum_{n \in \mathcal{N}_m} (w_1)_n^2 - \sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} (w_1)_i (w_1)_j.\end{aligned}$$

With $(w_1)_i = 0$, if $i \in \mathcal{N}_m$,

$$\lambda_1(A) \geq \lambda_1(A_m(\mathcal{N})).\tag{6.19}$$

From the inequality (6.18) and (6.19), we arrive at the bounds (6.13) of $\lambda_1(A_m(\mathcal{N}))$. \square

Theorem 6.4.1 implies that the decrease of spectral radius by removing a node or a set of nodes is strongly related to the principal eigenvector components corresponding to the removed nodes. Based on Theorem 6.4.1, we obtain a lower bound for components of the principal eigenvector

$$(x_1)_n \geq \sqrt{\frac{\lambda_1(A) - \lambda_1(A_1(\mathcal{N}))}{2\lambda_1(A)}}.\tag{6.20}$$

Motivated by Theorem 6.4.1, the eigenvector based one node removal strategy to minimize the largest eigenvalue simply removes the node with the largest principal eigenvector component $(x_1)_n$.

We perform further simulations to illustrate the importance of the principal eigenvector components in characterizing the influence of the link/node removal on λ_1 . We

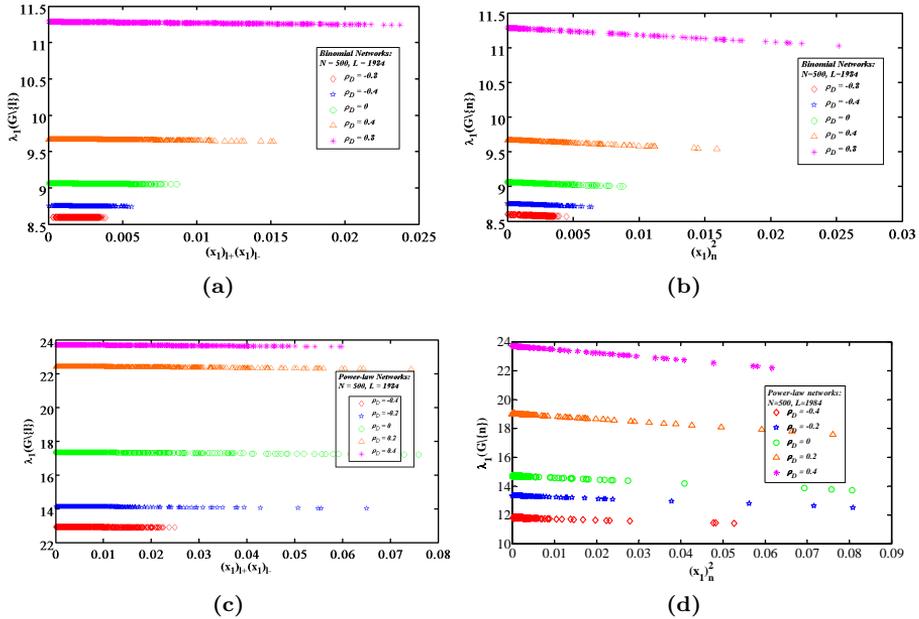


Figure 6.12: Spectral radius of graphs by removing a link (or node) as a function of corresponding components of the principal eigenvector (a), (b) in binomial graphs, (c), (d) in power-law graphs.

deduce networks with different assortativities but with a given degree vector, which may follow a binomial or power-law degree distribution. Upon each network, we try all possible one link (or node) removal and examine the largest eigenvalue $\lambda_1(G \setminus (l))$ (or $\lambda_1(G \setminus (n))$) after removing one link (or node) as a function of $(x_1)_{l^+}$ ($(x_1)_{l^-}$ (or $(x_1)_n^2$) corresponding to the link (or node) removed. By the Perron-Frobenius theorem [14], all components of x_1 and w_1 are non-negative (positive if the corresponding graph is connected). Interestingly, $\lambda_1(G \setminus (l))$ (or $\lambda_1(G \setminus (n))$) decreases linearly as a function of increasing $(x_1)_{l^+}$ ($(x_1)_{l^-}$ (or $(x_1)_n^2$), as shown in Figure 6.12. In other words, the spectral radius will be decreased more if the link (or node) removed has a larger $(x_1)_{l^+}$ ($(x_1)_{l^-}$ (or $(x_1)_n^2$).

The addition of a node to a graph G_N was discussed in [14, p. 60]. In particular, when G_{N+1} is the cone of a regular graph G_N , the spectral radius $\lambda_1(A_{N+1})$ of G_{N+1} equals $\frac{\lambda_1(A_N)}{2} \left(1 + \sqrt{1 + 4 \frac{d_n}{\lambda_1(A_N)^2}} \right)$, where $\lambda_1(A_N)$ is the spectral radius of G_N and $d_n = N$ is the degree of the added cone node. Hence, the increase of the spectral radius is related to the degree d_n . Lemma 6.4.1 shows that the decrease of the spectral radius by removing a node n is related to $(x_1)_n$ and complements the lemma 6.3.1 on link removals, proved in [108]. Lemma 6.3.1 relates the decrease of λ_1 by m link removals to the product $(x_1)_i(x_1)_j$. Moreover, the lower bound in (6.13) of the spectral radius by removing m nodes contains the term

$$\sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij}(x_1)_i(x_1)_j$$

illustrating that, if there are links between removed nodes (*i.e.* $l^+ = i$ and $l^- = j$),

the decrease of the spectral radius also depends on the product $(x_1)_i(x_1)_j$ over links corresponding to the connected nodes.

In addition, the upper bound in (6.13) of $\lambda_1(A_m(\mathcal{N}))$ states that the spectral radius λ_1 of a graph G is always larger than or equal to the largest eigenvalue of any subgraph G_s of G ,

$$\lambda_1 \geq \max_{\text{all } G_s \subset G} (\lambda_1(A_{G_s})),$$

which is another proof for Theorem 42 in [14, p. 246 – 247].

Goh *et al.* [143] observed by simulations in Barabasi-Albert graphs that the upper bound of $(x_1)_{max}^2$ is $\frac{1}{2}$, where $(x_1)_{max}$ is the largest component of the principal eigenvector. Corollary 6.4.2 provides a rigorous proof of this observation.

Corollary 6.4.2. *In any graph, any eigenvector component of the principal eigenvector obeys*

$$(x_1)_n \leq \frac{\sqrt{2}}{2} \tag{6.21}$$

Moreover,

$$\sum_{n \in \mathcal{N}_m} (x_1)_n^2 \leq \frac{1}{2} \left\{ 1 + \frac{1}{\lambda_1(A)} \sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} (x_1)_i (x_1)_j \right\} \tag{6.22}$$

Proof: Since all components of x_1 and $\widetilde{A}_1(\mathcal{N})$ are non-negative by the Perron-Frobenius Theorem, we have that $x_1^T (\widetilde{A}_1(\mathcal{N})) x_1 \geq 0$. Combining (6.15), (6.16) and $\lambda_1(A) > 0$, we obtain $(1 - 2(x_1)_n^2) \geq 0$, from which (6.21) follows. By the same argument $x_1^T (\widetilde{A}_m(\mathcal{N})) x_1 \geq 0$ and

$$\left(1 - 2 \sum_{n \in \mathcal{N}_m} (x_1)_n^2 \right) \lambda_1(A) + \sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} (x_1)_i (x_1)_j \geq 0$$

proving (6.22). □

Alternatively, the inequality in the proof also yields

$$\lambda_1(A) \geq \frac{\sum_{j \in \mathcal{N}_m} \sum_{i \in \mathcal{N}_m} a_{ij} (x_1)_i (x_1)_j}{2 \sum_{n \in \mathcal{N}_m} (x_1)_n^2 - 1} = \frac{\sum_{l \in \mathcal{L}_m^*} (x_1)_{l^+} (x_1)_{l^-}}{2 \sum_{n \in \mathcal{N}_m} (x_1)_n^2 - 1},$$

where \mathcal{L}_m^* denotes the set of links among the set \mathcal{N}_m of nodes removed from G . The sharpest bound is likely reached when $2 \sum_{n \in \mathcal{N}_m} (x_1)_n^2 \gtrsim 1$.

We remark that equality in (6.21) is reached for the star, when the node n is the central or hub node. Since scale-free graphs consists of few very high degree nodes, their influence on the eigenvector is close to a star, which explains the observations of Goh *et al.* [143]. When \mathcal{N}_m is an independent set (*i.e.* there are no links between the nodes of \mathcal{N}_m such that $a_{ij} = 0$ for any $i, j \in \mathcal{N}_m$), the non-negative double sum in (6.22) disappears and we find that

$$\sum_{n \in \mathcal{N}_m} (x_1)_n^2 \leq \frac{1}{2}.$$

This special case of (6.22) has been proved earlier by Cioaba [130].

Finally, the lower bound in (6.14) underlines the interpretation of a principal eigenvector component as an importance or centrality measure. For, the more important the node n is, the higher the value of $(x_1)_n$, and the larger the possible decrease in spectral radius when this node n is removed.

Recently, Felix Goldberg has found another lower bound for components of the principal eigenvector $(x_1)_n \geq 1/\sqrt{1 + \frac{d_n}{(\lambda_1(A) - \lambda_1(A_1(\mathcal{N})))^2}}$. His results show that his lower bound works better than ours (see Eq. (6.20)) for high-degree nodes, and our lower bound works for low-degree nodes.

6.5 Chapter Conclusions

The spectral radius is both fundamental in graph theory as well as in many dynamic processes in complex networks such as epidemic spreading, synchronization and reaching consensus [14, p. 200]. We have shown that the spectral radius minimization problem (for both link and node removals) is an NP-hard problem, which opens the race to find the best heuristic. In particular, in large infrastructures such as transportation networks, where removing links can be very costly, a near to optimal strategy is desirable. We have shown that an excellent strategy for link removal is $S = x_i x_j$. On average, this strategy outperforms most other heuristics, but it does not beat them at all times. Moreover, removing the node u with the highest vector component $(x_1)_u$ will decrease $\lambda_1(A)$ most. Beside graph theoretic bounds and arguments that underline the goodness of the heuristic $S = x_i x_j$, two scaling laws (6.9) and (6.12) are found: these laws may help to estimate the decrease in spectral radius as a function of the number N of nodes and/or the number m of link removals. It may be worthwhile that further investigations compute or estimate the scaling parameters β in (6.9) as well as α in (6.12).

CHAPTER 7

Increasing the Epidemic Threshold by Link Rewiring or Resetting

“There is no intelligence where there is no need of change.”

H.G. Wells, 1866 - 1946

EPIDEMICS have so far been mostly studied in undirected networks. However, many real-world networks, such as the online social network Twitter and the world-wide web, on which information, emotion or malware spreads, are directed networks, composed of both unidirectional links and bidirectional links. We define the *directionality* ξ as the percentage of unidirectional links in [70]. We propose two algorithms to generate directed networks with a given directionality ξ . The effect of ξ on the spectral radius λ_1 , principal eigenvector x_1 , spectral gap ($\lambda_1 - |\lambda_2|$) and algebraic connectivity μ_{N-1} is studied. Important findings are that the spectral radius λ_1 decreases with the directionality ξ , whereas the spectral gap and the algebraic connectivity increase with the directionality ξ . The extent of the decrease of the spectral radius depends on both the degree distribution and the degree-degree correlation (assortativity) ρ_D . In directed networks, the epidemic threshold is larger and a random walk converges to its steady-state faster than that in undirected networks with the same degree distribution.

7.1 Introduction

Much effort has been devoted to understand epidemics on networks, mainly because of the increasing threats from cybercrime and the expected outbreak of new fatal viruses in populations. Epidemics have been studied on undirected networks for a long time and many authors (see [79, 11, 118, 72, 119, 73, 31]) addressed the existence of an epidemic threshold τ_c in the susceptible-infected-susceptible (SIS) epidemic process[117]. Topologies of undirected networks have been mostly modeled by Erdős and Rényi [1, 144, 2] as

Table 7.1: *Percentage of unidirectional links in real-world networks*

Real-world networks	N	L_{arcs}	ξ
Enron	69,244	276,143	84.29%
Ljournal-2008	5,363,260	79,023,142	25.32%
Twitter-2010	41,652,230	1,468,365,182	64.29%
WordAssociation-2011	10,617	72,172	76.77%
cnr-2000	325,557	3,216,152	70.33%
in-2004	1,382,908	16,917,053	60.68%
eu-2005	862,664	1,935,140	67.80%
uk-2007-05@100000	100,000	3,050,615	82.23%
uk-2007-05@1000000	1,000,000	41,247,159	79.71%

binomial networks, by Bárábasi and Albert [7] as power-law networks, or by Watts and Strogatz [6] as small-world networks. More complicated static and dynamic models, such as the configuration model [145, 146, 114], are also proposed to approximate real-world networks. However, many real-world networks are *directed* networks, some examples are shown in Table 7.1. The dataset of the real-world networks is obtained from [147, 148] and the description of these networks is attached in the Appendix B.

Two kinds of links, namely bidirectional links and unidirectional links, exist in directed networks. If node i is connected to node j (denoted by $i \rightarrow j$) then j is also linked to i (denoted by $j \rightarrow i$), one bidirectional link exists between nodes i and j ; and if either $i \rightarrow j$ or $j \rightarrow i$ exists, but not both in between the node pair i and j , a unidirectional link exists. Here, we define the directionality as $\xi = L_{unidirectional}/L_{arcs}$, where the number of arcs (the number of nonzero elements in the adjacency matrix) $L_{arcs} = \sum_i \sum_j a_{ij} = u^T A u$, (u is the all-one vector), can also be calculated by $L_{arcs} = L_{unidirectional} + 2L_{bidirectional}$. A directed network with directionality ξ is denoted by $G^{(\xi)}$. The network $G^{(\xi=0)}$ is a bidirectional network or an undirected network, whose adjacency matrix is symmetric, when $\xi = 0$. The network $G^{(\xi=1)}$ is a directed network without any bidirectional link, when $\xi = 1$. A high directionality is observed in Twitter, as shown in Table 7.1. A link runs from user A to user B if user A follows user B in Twitter, where user A is called the "follower" of user B. The fact that user A "follows" user B, does not necessarily mean that the reverse is also true. For example, a famous person could have millions of followers but he/she may not follow many others. This explains the high directionality ξ of Twitter. The virtual-community social networks, such as LiveJournal, have a low directionality (see Table 7.1), mainly because they aim to construct virtual connections in between real-life friends, and friendship relations are usually mutual.

There has been an increasing interest in the study of directed networks. Topological properties of directed networks, such as the short loops, closure connectivity, degree, domination and communities on realistic directed networks have already been studied in [149, 150, 151, 152, 91, 153]. Garlaschelli and Loffredo [154] investigated the reciprocity [12] in directed networks, where the reciprocity is equal to $1 - \xi$. Processes taking place on networks, such as synchronization, percolation and epidemic spread, have also been researched [154, 155, 156, 71, 157] in real directed networks. Percolation theory for directed networks with $\xi = 1$ was firstly developed by Newman *et al.* [114, 158]. Then, Boguñá and Serrano [159] pointed out that even a small fraction of bidirectional links suffices to

percolate the network. Moreover, Meyers *et al.* [160] used a generating function method to predict the epidemic threshold in directed networks with $\xi < 1$ and the size of the infected cluster. Recently, Van Mieghem and van de Bovenkamp have proven that the NIMFA epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$ of the SIS epidemic process also holds for directed networks [161]. Stimulated by the directed social networks with different directionalities, here, we focus on the influence of the directionality ξ on the epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$ and other spectral properties.

This chapter is organized as follows. In Section 7.2, we propose two algorithms that could be applied to a bidirectional network to generate a directed network with an arbitrary given directionality ξ , by rewiring or resetting links. The in- and out-degree distribution of the generated directed network is the same as the degree distribution of the original bidirectional network. Chen and Olvera-Cravioto [162] proposed an algorithm to generate a directed network with a given in- and out-degree distribution, which is similar to the configuration model. However, their algorithm in [162] cannot generate a directed network with a given directionality ξ . In Section 7.3, we investigate the effect of the directionality ξ on the spectral radius λ_1 , the principal eigenvector x_1 (the eigenvector corresponding to λ_1), the spectral gap $\lambda_1 - |\lambda_2|$ and the algebraic connectivity μ_{N-1} in both directed binomial¹ and power-law networks, whose in-degree and out-degree both follow a binomial (or power-law) distribution. Interestingly, we find that the spectral radius λ_1 of networks $G^{(\xi=0)}$ is larger than that of directed networks $G^{(\xi=1)}$ when the degree distribution and the assortativity of these networks are the same. This means that the epidemic threshold τ_c in undirected networks is smaller than that in directed networks with the same degree distribution and assortativity. Furthermore, we explore the influence of the Pearson degree correlation coefficient ρ_D (also called the assortativity) on the epidemic threshold $\tau_c^{(1)}$ in both directed binomial and directed power-law networks with different ξ , in Section 7.4. The ρ_D is the Pearson correlation coefficient of degrees [17] at either ends of a link and lies in the range $[-1, 1]$. Actually, there are four degree correlations, namely the in-degree and in-degree correlation, the in-degree and out-degree correlation, the out-degree and in-degree correlation and the out-degree and out-degree correlation, in directed networks. We consider directed networks where the in-degree and out-degree of each node are the same. In this case, the four degree correlations are equal to each other and can be all referred as the degree correlation (or the assortativity). The decrease of the spectral radius λ_1 with ξ is large when the assortativity ρ_D is large in directed binomial networks, whereas the opposite is observed in directed power-law networks.

7.2 Algorithm Description

Here, we propose two algorithms, In-degree and Out-degree Preserving Rewiring Algorithm (IOPRA) and Link resetting algorithm (LRA), which both can be applied to any network to generate a directed network with a given directionality ξ . In this study, we only apply these two algorithms to generate directed networks with the same in- and out-degree distribution. The difference is that IOPRA preserves the in- and out-degree of

¹For example, an Erdős-Rényi random network is a binomial network with the Pearson degree correlation $\rho_D = 0$. A general binomial network could possibly have an assortativity ρ_D within a large range.

each node, while, LRA may change the in- and out-degree of any node. IOPRA is inspired by the degree preserving rewiring, which has been presented in [40, 163]. We firstly introduce the degree-preserving rewiring, which monotonously increases or decreases the assortativity ρ_D , while maintaining the node degrees unchanged, in undirected networks. Afterwards, we describe IOPRA and LRA in detail.

7.2.1 Degree-preserving Rewiring

The degree-preserving rewiring [40] can either increase or decrease the assortativity of a bidirectional network: (a) the degree-preserving assortative rewiring: randomly select two links associated with four nodes and then rewire the two links such that the two nodes with the highest degree and the two lowest-degree nodes are connected, respectively. If any of the newly rewired links exists before rewiring, discard this step and a new pair of links is randomly selected; (b) the degree-preserving disassortative rewiring: randomly select two links associated with four nodes and then rewire the two links such that the highest-degree node and the lowest-degree node are connected, and the remaining two nodes are also connected, as long as the newly rewired links do not exist before rewiring. Either rewiring step (a) or (b) can be repeated to monotonically increase or decrease the assortativity in a bidirectional network.

7.2.2 In-degree and Out-degree Preserving Rewiring Algorithm (IOPRA)

IOPRA can be applied to change the directionality of networks. We define our In-degree and Out-degree Preserving Rewiring Algorithm (IOPRA) as follows: randomly choose two unidirectional links with four end nodes, and rewire the two unidirectional links. In IOPRA, the head of one unidirectional link only can rewire with the head of the other unidirectional link, in order to maintain both the in-degree and out-degree of the four nodes unchanged (see Figure 7.1). We don't rewire if such rewiring can introduce duplicated links from any node to any other. We discard the rewiring step if this rewiring step doesn't change the directionality ξ towards the given directionality. In both cases, we randomly reelect a pair of unidirectional links associated with four nodes. We illustrate the process of IOPRA changing the directionality in Algorithm 1 (see Appendix D). IOPRA actually changes the directionality ξ of a given network G without changing the in- and out-degree of each node. If the original network G is an undirected network, the in-degree sequence is exactly the same as the out-degree sequence in the directed network $G^{(\xi)}$ generated by IOPRA.

IOPRA changes the directionality, as well as randomizing the connections of the original network, without changing the degree of any node. Hence, if the initial network is a random network, e.g. an Erdős-Rényi (ER) network or a scale-free network generated by the configuration model, where the connections are originally laid at random, IOPRA changes only the directionality ξ . However, if we apply IOPRA to a nonrandom network, e.g. a lattice, the resulting network has not only a different directionality but also a more randomized structure.

7.2.3 Link Resetting Algorithm (LRA)

We start with a bidirectional network G , and use the Link Resetting Algorithm (LRA) to change the directionality (see Algorithm 2 in Appendix D). We randomly choose a fraction

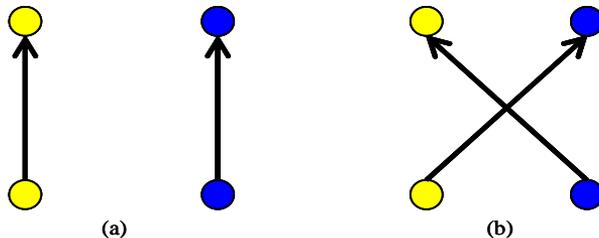


Figure 7.1: *In-degree and out-degree preserving rewiring*

ξ of the bidirectional link pairs from G . Then, we randomly choose only one unidirectional link from each bidirectional link, and randomly relocate the selected unidirectional links to a place without any link. In this work, we only apply LRA to ER networks. In this case, the in- and out-degree of the generated directed network follow the same binomial degree distribution as the original network. However, the in-degree and out-degree of any node in the generated network may differ from those in the original network G . When LRA is applied to other types of networks, such as the power-law networks, the original in- and out-degree distributions are destroyed, and tend to be binomial.

In summary, two types of directed binomial networks can be generated: one is generated by IOPRA (called the IOPRA directed binomial networks), whose nodes have the same in-degree and out-degree; the other, created by LRA (called the LRA directed binomial networks), has the same in- and out-degree distribution, while allowing the in- and out- degree of any node to be different.

7.3 Spectral Properties in Directed Networks

7.3.1 Spectral Radius of Directed Networks

The adjacency matrix of a directed network is an asymmetric matrix, whose spectral radius λ_1 is still real by the Perron-Frobenius Theorem (see [14]). We generate directed networks, with the directionality ξ ranging from 0 to 1, by applying IOPRA to the ER ($N = 1000$, $p = 2\ln N/N$) and the BA ($N = 1000$, $m = 4$) networks gradually. Here we choose $p \geq 2p_c$ to be sure that the original networks are connected. The influence of the directionality ξ on the spectral radius λ_1 and the assortativity ρ_D is studied in both directed power-law networks and directed binomial networks (see Figure 7.2).

Apart from some wobbles, the spectral radius λ_1 decreases almost linearly with the directionality ξ . The same phenomenon can also be observed in large, sparse directed networks [70]. Moreover, the assortativity ρ_D of the network fluctuates slightly around 0. We also have observed a similar phenomenon in large sparse networks. We observe that the tiny leaps of spectral radius λ_1 happen when the assortativity ρ_D has a rise, which is understandable, because it has been shown in [40] that the spectral radius λ_1 increases with the increase of the assortativity ρ_D . Figure 7.3 exemplifies that the spectral radius λ_1 may increase instead of decreasing when the directionality increases due to the assortativity ρ_D . We will study the effect of the assortativity ρ_D on the decrease of the

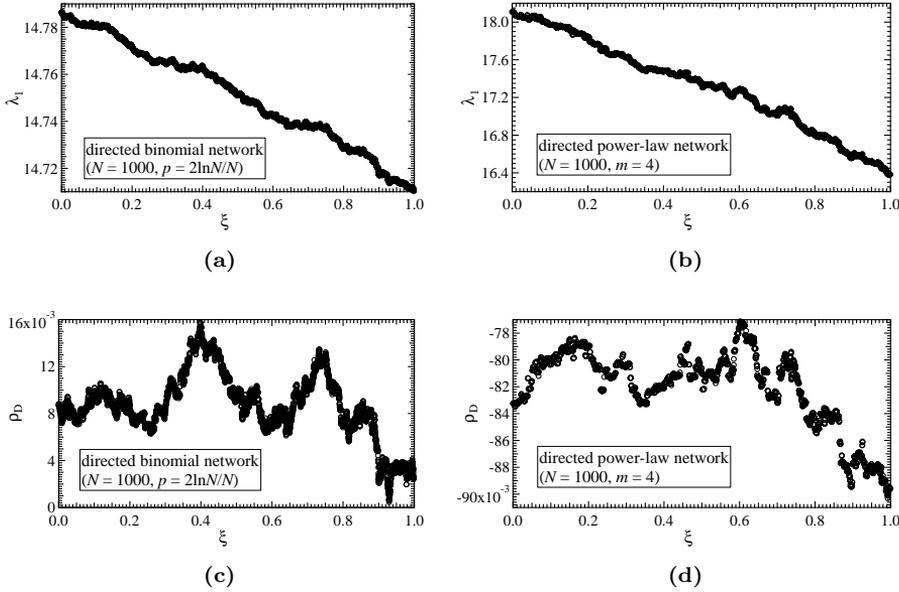


Figure 7.2: Plot of the spectral radius versus the directionality in (a) and (b), as well as the assortativity versus the directionality in (c) and (d), in both directed binomial and power-law networks generated by IOPRA.

spectral radius λ_1 with the directionality ξ in Section 7.4.

With LRA, we generate directed binomial networks with directionality ξ from 0 to 1 with step 0.1. The assortativity ρ_D of all the directed binomial networks generated by LRA is around 0. Hence, the effect of the assortativity ρ_D can be ignored here. The spectral radius λ_1 is calculated in directed networks with different directionality ξ . We performed all the simulations for 10^3 network realizations. The spectral radius λ_1 is plotted as a function of the directionality ξ for directed binomial networks with $p = 2\ln N/N$ and $p = 0.05$ in Figure 7.4. From the observation, the spectral radius λ_1 is inversely proportional to the directionality ξ with the factor $\simeq -1$, which is independent from the link density p of the networks. This observation can be explained by the following proposition.

Proposition 7.3.1. *Let $G^{(\xi=0)} = G_p(N)$ be a connected Erdős-Rényi (ER) random graph with a finite N , and let $G^{(\xi)}$ be a directed binomial network generated by LRA whose in- and out-degree follow the same binomial distribution as $G_p(N)$. The average spectral radius satisfies*

$$E[\lambda_1(G^{(\xi)})] \simeq E[\lambda_1(G_p(N))] - \xi \quad (7.1)$$

Arguments:

A directed binomial network $G^{(\xi)}$ generated by LRA with link density p , can be equivalently constructed by randomly adding $2p\xi\binom{N}{2}$ unidirectional links to a bidirectional ER network $G_{p(1-\xi)}(N)$ with size N and link density $p(1-\xi)$. The average spectral radius

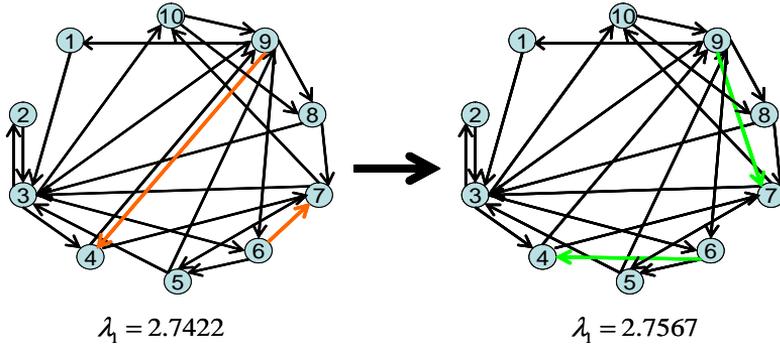


Figure 7.3: Example: the spectral radius increases with the directionality ξ , because of the increase of the assortativity (where $\rho_D(G_{left}) = -0.6190$, $\xi(G_{left}) = 0.8333$, and $\rho_D(G_{right}) = -0.5714$, $\xi(G_{right}) = 0.9167$).

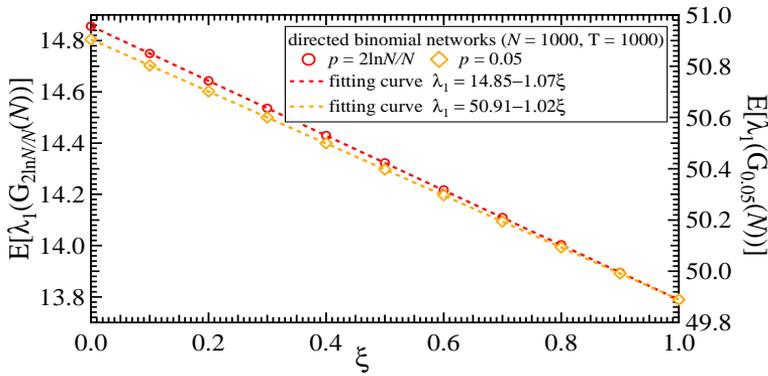


Figure 7.4: Average spectral radius as a function of the directionality for directed binomial networks generated by LRA with size $N = 1000$. Two values for the link density p are shown: $p = 2\ln N/N$ (red circles) and $p = 0.05$ (orange diamonds).

[14, pp. 173, art. 137] of $G_{p(1-\xi)}(N)$ is

$$E[\lambda_1(G_{p(1-\xi)}(N))] = (N-2)p(1-\xi) + 1 + O\left(\frac{1}{\sqrt{N}}\right).$$

The principal eigenvector of an adjacency matrix A is denoted by x_1 obeying the normalization $x_1^T x_1 = 1$. Let C denote the adjacency matrix of the resulting network after adding one unidirectional link to network G . The largest eigenvalue is increased due to the addition of the link ($i \rightarrow j$) [14, pp. 236, Lemma 7] as

$$\lambda_1(C) \simeq \lambda_1(A) + (x_1)_i (x_1)_j,$$

where the increase is strict if the adjacency matrix A is irreducible. Hence, the average increase of the spectral radius by adding m unidirectional links in random networks is obtained as

$$E[\lambda_1(C) - \lambda_1(A)] \simeq mE[(x_1)_i (x_1)_j].$$

The sum of the product of components in the principal eigenvector of Erdős-Rényi networks is approximated by a function of link density p (see Figure 7.5). The fitting function can be expressed as,

$$E \left[\sum_{j=1}^N \sum_{i=1}^N (x_1)_i (x_1)_j \right] = N - \frac{1}{p} + O(1),$$

when the network is connected. Since $x_1^T x_1 = 1$ and since the expectation $E[\cdot]$ is a linear operator, we obtain

$$E [(x_1)_i (x_1)_j] = \frac{N - \frac{1}{p} - 1}{N(N - 1)} + O\left(\frac{1}{N^2}\right), \tag{7.2}$$

when $i \neq j$. Directed binomial networks generated by LRA from ER with N and p , have the same $E [(x_1)_i (x_1)_j]$. Hence, the average spectral radius of the directed network obtained by adding $m = 2p\xi\binom{N}{2}$ unidirectional links to the network $G_{p(1-\xi)}(N)$ can be approximated by

$$E[\lambda_1(G^{(\xi)})] \simeq E[\lambda_1(G_{p(1-\xi)}(N))] + 2(N(N - 1)/2)p\xi E [(x_1)_i (x_1)_j].$$

Using (7.2),

$$E[\lambda_1(G^{(\xi)})] \simeq (N - 2)p + 1 - \xi + O\left(\frac{1}{\sqrt{N}}\right),$$

which leads to (7.1). □

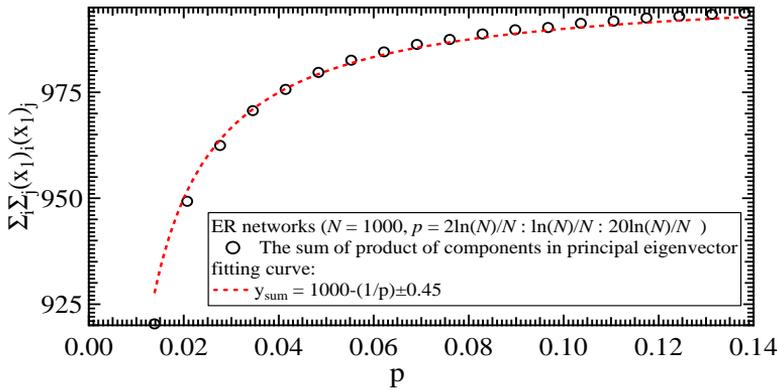


Figure 7.5: Sum of the product of components in the principal eigenvector as a function of the link density p in ER networks ($N = 1000$).

Juhász [164] also pointed out that the largest eigenvalue $\lambda_1(G^{(\xi=1)})$ of a directed random network with link density p and size N is almost surely Np , when N is large. In ER random networks, the spectral radius $E[\lambda_1(G^{(\xi=0)})] \rightarrow Np + 1$, when N is large (see [14, pp. 173, art. 137]). Both earlier results are consistent with Proposition 7.3.1, and support that the proportionality factor between the spectral radius λ_1 and the directionality ξ is around -1 .

Proposition 7.3.1 also reveals the effect of the size N on the relative largest decrease of the spectral radius $\Lambda = \frac{\lambda_1(G^{(\xi=0)}) - \lambda_1(G^{(\xi=1)})}{\lambda_1(G^{(\xi=0)})}$. We predict that $\Lambda \rightarrow 0$ if $N \rightarrow \infty$ for directed binomial networks, because the decrease of the spectral radius ($\lambda_1(G^{(\xi=0)}) - \lambda_1(G^{(\xi=1)})$) is almost a constant value, whereas the spectral radius $\lambda_1(G^{(\xi=0)})$ of dense directed binomial networks increases with the size of the networks. This implies that the effect of the directionality ξ on the spectral radius is small in large dense binomial networks.

7.3.2 Principal Eigenvector in Directed Networks

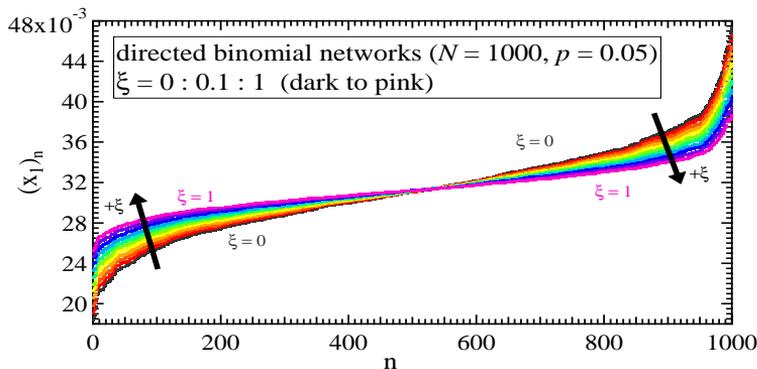


Figure 7.6: Change of the components of principal eigenvector from bidirectional networks to directed networks.

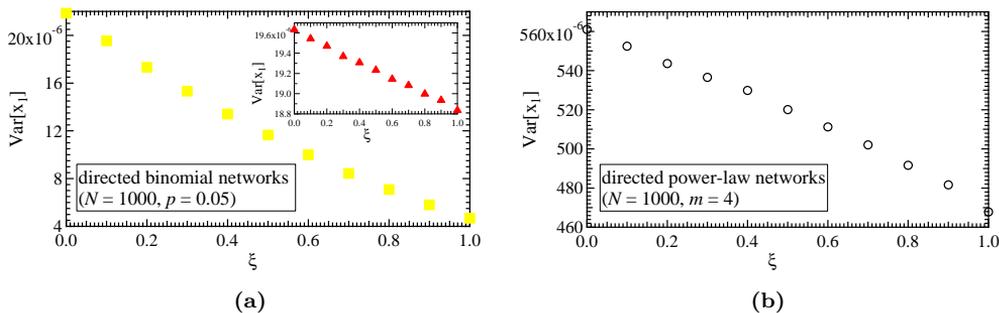


Figure 7.7: Plot of the variance of the principal eigenvector versus the directionality (a) in directed binomial networks (generated by LRA in yellow squares and by IOPRA in red triangles) and (b) in directed power-law networks by IOPRA (10^3 network realizations)

The principal eigenvector x_1 was first proposed as a centrality metric by Bonacich [165] in 1987, to indicate the influence of each node. For example, the decrease of the spectral radius [108, 129] by removing nodes, can be characterized by the corresponding principal eigenvector components. In this section, we explore the principal eigenvector in directed

networks. The principal eigenvector of the directed networks, with the directionality ξ from 0 to 1 with step 0.1, are calculated. Then, the components of the principal eigenvector are sorted in an ascending order. For each ξ , we simulate 10^3 network realizations and compute the average sorted principal eigenvector components. Figure 7.6 illustrates that the components of the principal eigenvector are more uniform in directed binomial networks: the principal eigenvector $x_1 \rightarrow \frac{u}{\sqrt{N}}$ as $\xi \rightarrow 1$; moreover, the variance of components of the principal eigenvector linearly decreases with the directionality ξ in both directed binomial networks and the directed power-law networks (see Figure 7.7). The observation implies that when the directionality is larger, the influence of each node on the spectral radius is more similar. This experimental evidence suggests that increasing the directionality enables all nodes to contribute more similarly to the robustness against epidemic in directed networks.

The decrease of the variance $Var[x_1]$ in directed binomial networks by LRA is larger than that in directed binomial networks by IOPRA (see Figure 7.7(a)). The connections in LRA directed binomial networks are more random than that in IOPRA directed binomial networks, in the sense that LRA allows each node to have a different in- and out- degree, although the in- and out- degree distribution are the same in both LRA and IOPRA binomial networks. As a consequence, the principal eigenvector x_1 is more uniform with a smaller $Var[x_1]$ in LRA directed binomial networks than that in IOPRA directed binomial networks when the directionality is the same. Thus, nodes in LRA directed binomial networks have more equal contributions to the spectral radius than nodes in IOPRA directed binomial networks with the same directionality. Li *et al.* [107] have shown that both a large variance of the degree and a large assortativity ρ_D contribute to a large variance $Var[x_1]$ of the components of the principal eigenvector x_1 . Here, we point out further that a large directionality ξ leads to a small variance $Var[x_1]$ of the components of x_1 .

7.3.3 Spectral Gap of Directed Networks

The difference $(\lambda_1 - \lambda_2)$ between the largest eigenvalue λ_1 and the second largest eigenvalue λ_2 is called the spectral gap. All eigenvalues of the symmetric adjacency matrix of an undirected network are real. Here we focus on the directed networks, whose adjacency matrix is asymmetric. The eigenvalues of directed networks can be complex numbers (as exemplified in Figure E.1 in Appendix E). In directed networks, the spectral gap is defined as $\lambda_1 - |\lambda_2|$, where $|\lambda_2|$ is the modulus of λ_2 . The spectral gap $\lambda_1 - |\lambda_2|$ increases with the directionality ξ in both the directed binomial networks and the directed power-law networks (see Figure 7.8). As introduced in Sec. 7.3.1, the spectral radius decreases with the directionality. Our observation implies that the second largest eigenvalue $|\lambda_2|$ decreases with the directionality faster than the spectral radius. The larger the spectral gap is, the faster a random walk converges to its steady-state [14, pp. 64]. Thus, the dynamic process in a directed network reaches the steady-state faster than that in an undirected network with the same degree distribution. Figure 7.8 (a) implies that a dynamic process is slightly faster to reach the steady-state in IOPRA directed binomial networks than in LRA directed binomial networks. The existence of large spectral gap together with a uniform degree distribution results in higher structural sturdiness and robustness against node and link failures [166]. Hence, directed networks with high directionality ξ and a uniform degree distribution are more robust than undirected networks with large

variance of degree.

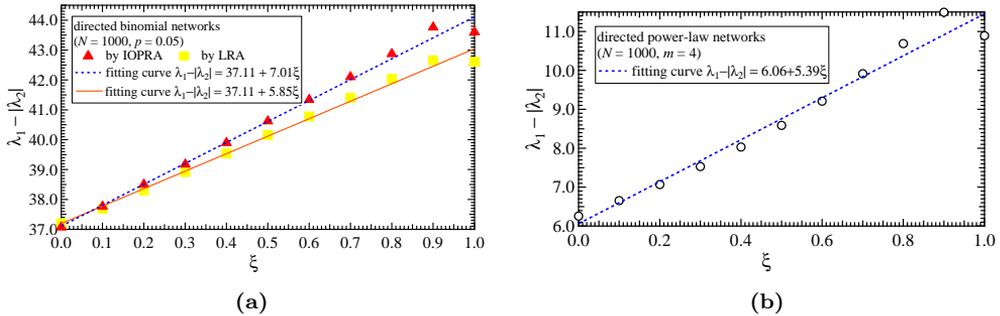


Figure 7.8: Plot of the spectral gap as a function of the directionality (a) in directed binomial networks (generated by LRA in yellow square and by IOPRA in red triangle) and (b) in directed power-law networks (10^3 network realizations).

7.3.4 Algebraic Connectivity of Directed Networks

The Laplacian matrix [167] is defined as $Q = \frac{1}{2}BB^T$, where the incidence matrix B is an $N \times L$ matrix with elements [14]

$$b_{il} = \begin{cases} 1 & \text{if link } e_l = i \rightarrow j \\ -1 & \text{if link } e_l = j \rightarrow i \\ 0 & \text{otherwise.} \end{cases}$$

The Laplacian matrix can be equivalently expressed as $Q = \Delta - \bar{A}$, where $\Delta = \frac{1}{2}(\Delta_{in} + \Delta_{out})$, Δ_{in} and Δ_{out} are diagonal matrices which contain the in-degree and out-degree of each node respectively, and $\bar{A} = \frac{1}{2}(A + A^T)$. If the network is an undirected network, \bar{A} is the adjacency matrix A and $\Delta = \text{diag}(d_1, d_2, \dots, d_N)$ is the degree matrix. The second smallest eigenvalue μ_{N-1} of the Laplacian Q was named algebraic connectivity by Fiedler [47]. The Laplacian Q is always symmetric as defined. Hence, the algebraic connectivity of a directed and connected network is a positive real number. The algebraic connectivity, together with the spectral gap, quantifies the robustness and the network's well-connectedness. The larger the algebraic connectivity is, the more difficult it is to cut the network into disconnected parts. Here, we study the influence of the directionality ξ on the algebraic connectivity μ_{N-1} of directed networks. As illustrated in Figure 7.9, the algebraic connectivity increases with the directionality ξ in both the directed binomial networks and the directed power-law networks. This suggests that the directed networks with high directionality are more difficult to break into parts and synchronize faster. As the directionality increases, the number of non-zero elements of \bar{A} increases and the variance of the elements of \bar{A} decreases. This could be one possible reason why the network is better connected. Moreover, the algebraic connectivity μ_{N-1} is greater in the LRA directed binomial networks than in the IOPRA binomial networks (see Figure 7.9(a)).

The algebraic connectivity μ_{N-1} approaches the spectral gap $\lambda_1 - \lambda_2$, as the network tends to be regular bidirectional networks [14, pp. 71], which suggests the spectral gap is related to the algebraic connectivity. Figures 7.8 and 7.9 show that both, the spectral

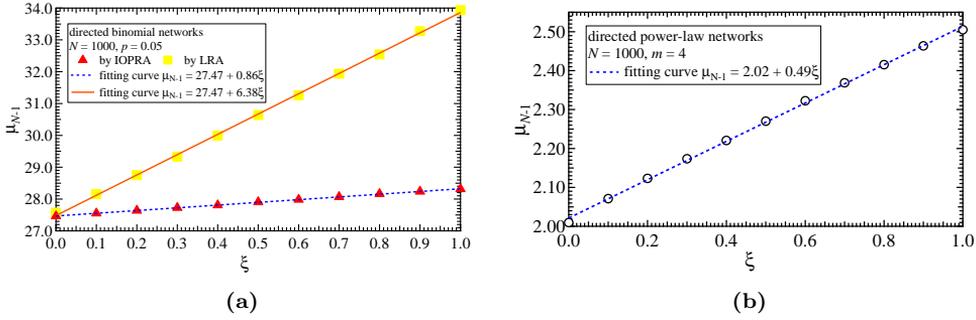


Figure 7.9: Plot of the algebraic connectivity as a function of the directionality (a) in directed binomial networks (generated by LRA in yellow square and by IOPRA in red triangle) and (b) in directed power-law networks (10^3 network realizations).

gap and the algebraic connectivity, increase with the directionality ξ in directed networks, which is consistent with the relation between the algebraic connectivity and the spectral gap.

7.4 Effects of the Assortativity on λ_1 of Directed Networks

In the directed networks generated by applying IOPRA to ER or BA networks, the in- and in-degree correlation, the in- and out-degree correlation, the out- and in-degree correlation and the out- and out-degree correlation are the same. Thus, the four correlations are all referred as the degree correlation (or the assortativity). In Section 7.3, we have discussed how the spectral properties change with the directionality in directed networks, where the assortativity is always close to zero. Here, we study how the spectral radius λ_1 changes with the directionality ξ when the assortativity ρ_D is the same, and how the change of the spectral radius λ_1 with ξ is influenced by the assortativity in directed networks. Two approaches are applied to investigate this problem.

- **Approach 1:** First, we perform degree-preserving rewiring on ER networks (or BA networks) to obtain a set of bidirectional networks with assortativity ρ_D from -0.8 to 0.8 (or -0.3 to 0.3) with step 0.1 . Second, we alter the directionality ξ of all bidirectional networks with each assortativity using IOPRA. The directionality ξ is changed from 0 to 1 with step 0.1 . IOPRA randomizes network connections, and thus pushes the assortativity of the resulting directed network towards zero, if the original network has a non-zero assortativity. Figure 7.10 plots the simulation results of one binomial network realization and 10^2 binomial network realizations. The simulation of one realization is almost the same as the result of a large number of network realizations, which points to *almost sure behavior* [168]. The results in the directed power-law networks are shown in Figure 7.11.

- **Approach 2:** First, we generate ER networks (or BA networks) $G^{(\xi=0)}$ whose directionality $\xi = 0$. Second, we apply IOPRA to ER (or BA) networks $G^{(\xi=0)}$ to generate directed binomial networks (or directed power-law networks) $G^{(\xi=1)}$ with directionality $\xi = 1$. Then, we change the assortativity of ER networks (or BA networks) $G^{(\xi=0)}$ and directed binomial networks (or directed power-law networks) $G^{(\xi=1)}$ by degree preserving

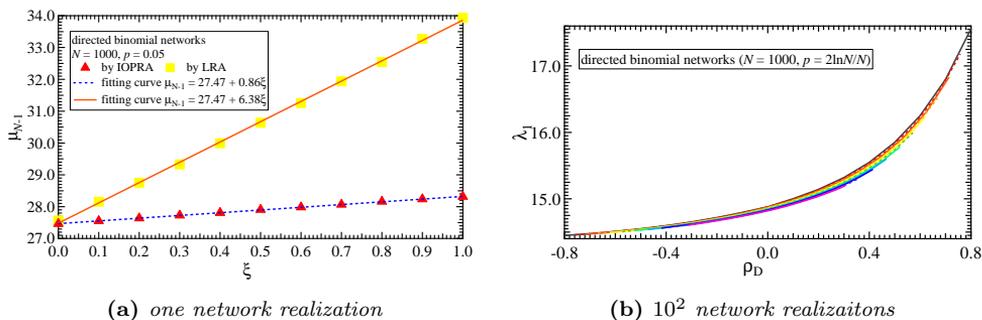


Figure 7.10: Spectral radius as a function of the assortativity in directed binomial networks with ξ from 0 to 1 with step 0.1 is scatter plotted in different colorful (or grayscale) lines, from the gray (upper) line to the pink (lower) line.

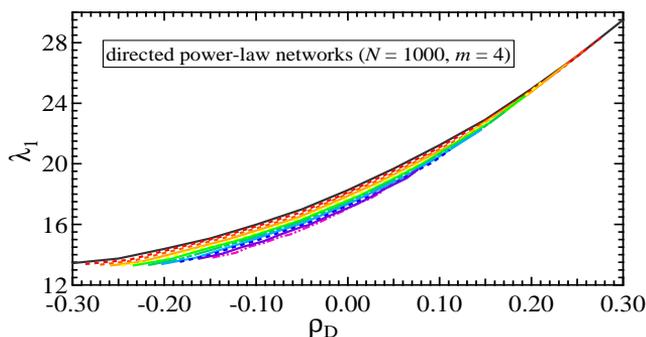


Figure 7.11: Spectral radius as a function of the assortativity in directed power-law networks with ξ from 0 to 1 with step 0.1, is scatter plotted in different colorful (or grayscale) lines, from the gray (upper) line to the pink (lower) line.

rewiring and in- and out-degree preserving rewiring, respectively, without changing the directionality. Note that the in- and out-degree preserving rewiring can be applied not only to change the directionality, but also to change the assortativity. Figure 7.12 plots the spectral radius of the networks $G^{(\xi=0)}$ and $G^{(\xi=1)}$ as a function of the assortativity. The two approaches both change the assortativity and the directionality of the networks, while, the order of change is different: Approach 1 changes the assortativity firstly and then the directionality; Approach 2 is the opposite. Figures 7.10, 7.11 and 7.12, show that the spectral radius λ_1 always decreases with the directionality ξ when the networks have the same degree distribution and the same assortativity ρ_D . Moreover, the degree distribution of the network also influences the change range of the spectral radius λ_1 with ξ . The decrement of the spectral radius λ_1 with ξ increases with the assortativity in directed binomial networks (see Figures 7.10 and 7.12 (a)). On the contrary, the decrement of the spectral radius λ_1 with ξ goes down with the assortativity in directed power-law networks (see Figures 7.11 and 7.12 (b)). Furthermore, the decrease of the spectral radius in directed power-law networks is larger than that in directed binomial networks, when the assortativity is zero. Many real-world networks are directed power-

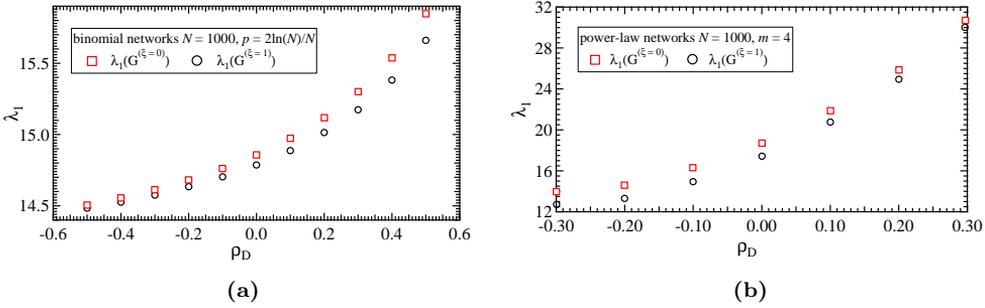


Figure 7.12: Spectral radius as a function of the assortativity (a) in directed binomial networks ($N = 1000$, $p = 2\ln N/N$) and (b) in directed power-law networks ($N = 1000$, $m = 4$) for 10^3 network realizations.

law networks, where λ_1 could possibly be tuned within a large range by controlling the directionality in real-world networks.

Summarizing, the spectral radius λ_1 decreases with the directionality ξ when the assortativity remains constant. In order to protect the network from virus spreading via increasing the epidemic threshold, while maintaining the degree distribution and the assortativity, increasing the directionality of networks is recommended. Meanwhile, the spectral gap and the algebraic connectivity are also increased, which means that the topological robustness is also enhanced in return.

7.5 Chapter Conclusions

In this chapter, two algorithms to generate directed networks with a given directionality ξ are proposed. This allows us to study the influence of the directionality ξ on the spectral properties of networks. The spectral radius λ_1 , which is the inverse of the SIS NIMFA epidemic threshold $\tau_c^{(1)}$, is studied in directed networks. A universal observation is that, the spectral radius decreases with the directionality when the degree distribution and the assortativity of the network is preserved. We may, thus, increase the epidemic threshold to suppress the virus spread via increasing the directionality of the network. The possible range to increase the epidemic threshold is relatively large in directed binomial networks with a high assortativity and directed power-law networks with a low assortativity. The variance of the components of the principal eigenvector decreases with the directionality, which indicates that the influence of each node on the spectral radius is similar in networks with a high directionality. Moreover, the spectral gap and the algebraic connectivity increase with the directionality, implying that an increase of the directionality enhances the connectivity of the network. Furthermore, we observe that the spectral gap increases faster with the directionality in IOPRA than in LRA directed binomial networks, on the contrary, the algebraic connectivity increases with the directionality faster in LRA than in IOPRA directed binomial networks. This observation may be due to the fact that the in- and out- degree of each node could be different in LRA directed binomial networks, while, are exactly the same in IOPRA directed binomial networks.

CHAPTER 8

Conclusions and Future Work

“Learn from yesterday, live for today, hope for tomorrow. The important thing is to not stop questioning.”

Albert Einstein, 1879 - 1955

8.1 Main Conclusions

The objective of researching complex networks is to provide deep insights into complex systems in the real world. The real complex systems are abstracted as networks composed of nodes and links. Network measurements are essential for many network investigations, including representation, characterization, classification and modeling. In recent decades, a number of metrics have been proposed to quantify the properties of networks. This thesis first provides a comprehensive review of the main measurements, including the structural metrics, the spectral metrics and the centrality metrics, in Chapter 2. However, we have no idea if the entire set of metrics is needed to characterize a network. We expect to find a small set of representative metrics, which are not redundant but enough for characterizing complex networks. The first part of this thesis is devoted to explore the relationships between metrics. From the theoretical analysis and simulations, we learn that there are correlations between certain network metrics and a network with a given degree distribution can indeed be characterized by a small representative set of structural and spectral metrics. We find that spectral properties are essential in network characterizations. What is more, we could approximate a high-complexity centrality metric by a strongly correlated low-complexity one. We gain new insight into the metrics/properties that are applied to capture the functions of complex networks. The functions of a network are usually expressed through the dynamic processes on the network. The notable successes in this area so far have been studies of the virus spreading over networks. The epidemic thresholds of SIS model can be expressed by metrics, which are studied and compared in this thesis. The second objective of this thesis is to design a network. We

change the topological structure of networks to reinforce certain functions (such as robustness, epidemic threshold and connectivity) of the networks, via three approaches, *i.e.* the link removal, the node removal and the directionality changing. We summarize the main findings of the research performed on the subject of characterization and design of complex networks as follows:

Chapter 3 investigates the Pearson correlation coefficients between widely studied structural and spectral network metrics in three network models (Bárabasi-Albert graphs, Erdős-Rényi random graphs and Watts-Strogatz small-world graphs) as well as in functional brain networks of healthy subjects. We find that the correlations between metrics are topology dependent. Each class of networks can be characterized by a metric correlation pattern, which illustrates the strong correlations and independencies between metrics. The high correlation between a pair of metrics indicates that there are large redundancies. Accordingly, we can determine a small set of representative metrics by including only one metric from each subset of mutually strongly dependent metrics, to characterize both sufficiently and effectively a class of networks with a given degree distribution. Note that the representative set at least has to be considered so that important network properties will not be overlooked, when we study a specific network. Most of the metric correlations observed so far are supported/explained analytically by theorems developed in chapter 3 as well as in the literature. Moreover, we observed that the unweighted networks, which are obtained from weighted functional brain networks with a fixed threshold, and Erdős-Rényi random graphs follow a similar degree distribution. This finding verifies the influence of the degree distribution on metric relations. This representative set of metrics tends to both sufficiently and effectively characterize networks with a given degree distribution. In the study of a specific network, however, we have to at least consider the representative set so that important network properties will not be neglected. Furthermore, two important conclusions for the applications of network science to other disciplines are drawn as: 1) the average distance and the clustering coefficient, the most studied metrics so far in neuroscience, are strongly correlated, thus, redundant; 2) spectral metrics, though only studied recently in the context of complex networks seem to be essential in network characterizations.

Chapter 4 studies the centrality metrics and their applications to the opinion model and the link removal. First, we explore the properties of the principal eigenvector x_1 , and find that: i) the average principal eigenvector $E[x_1]$ decreases with the increase of assortativity ρ_D ; ii) the upper bound of $E[x_1]$ is $1/\sqrt{N}$; iii) the component of the principal eigenvector increases exponentially, when ρ_D is large. Second, the relation between the principal eigenvector x_1 and the degree vector d is studied. We prove that the difference between the principal eigenvector x_1 and the degree vector d is the smallest when $\lambda_1 = N_2/N_1$ and also find that x_1 and d are strongly linear correlated when the assortativity is not small. Third, we introduce a novel centrality metric, the m th-order degree mass, which is defined as the sum of the weighted degree of the node and its neighbors not further away from m hops. The Pearson correlation coefficient and centrality similarity are applied to study the correlations between centrality metrics. We find that the betweenness, the closeness and the components of the principal eigenvector are strongly correlated with the degree, the 1st-order degree mass and the 2nd-order degree mass, respectively, in both network models and real-world networks. We theoretically prove that the linear correlation between the 2nd-order degree mass and the principal eigenvector is stronger than that between the principal eigenvector and the degree. In addition, we investigate

the effect of the selected inflexible contrarians based on different centrality metrics in helping one opinion to compete with the other in the inflexible contrarian opinion (ICO) models. We observe that selecting the inflexible contrarians based on the leverage, the betweenness or the degree is the most effective strategy in opinion-competition in all types of networks. The strongly correlated centrality metrics perform similarly to each other in applications, *i.e.* link removal and opinion model.

Chapter 5 compares two SIS mean-field approximations, the HMF approximation and the N-intertwined approximation, with the ε -SIS as a benchmark, in different graph types. We define the steady-state of ε -SIS spreading model as the meta-stable state of SIS model. The steady-state fraction of infected nodes in the ε -SIS model is the simplest and best way to determine the number of infected nodes in the meta-stable state of the SIS model. We compare the epidemic threshold and the steady-state fraction of infected nodes of the two approximations. The epidemic thresholds of the N-intertwined approximation and the HMF approximation are $\tau_c^{(1)} = 1/\lambda_1$ and $\tau_c^{\text{HMF}} = E[D]/E[D^2]$, respectively. In most cases, our simulations show that $\tau_c^{(1)} > \tau_c^{\text{HMF}}$, which implies that the $\tau_c^{(1)}$ is a better lower bound for SIS model. We also see that the N-intertwined approximation can approach the ε -SIS epidemic model well in most graph types. Overall, the N-intertwined approximation is superior to the HMF approximation. The N-intertwined approximation is exactly the same as the HMF approximation in regular graphs. However, for some special graph types, such as the square lattice graph and path graph, the two mean-field approximations are both far away from the ε -SIS spreading model.

Chapter 6 designs networks in order to enlarge the network's epidemic threshold τ_c , or, equivalently, to lower $\lambda_1(A)$. We perform two approaches, removing links and removing nodes, to make networks becoming less vulnerable to a virus. We prove that the minimization of the spectral radius by removing m links (or nodes) is shown to be an NP-complete problem. We demonstrate that removing the node u or the link $u \sim v$ with highest vector component $(x_1)_u$ or highest vector component product $(x_1)_u(x_1)_v$ will decrease $\lambda_1(A)$ most. We compare several greedy strategies for removing m links and observe that the strategy that removes that link $l = i \sim j$ with largest product $(x_1)_i(x_1)_j$ of the components of the eigenvector x_1 belonging to the largest adjacency eigenvalue is superior to other strategies in most cases. We find two scaling laws, which may help to estimate the decrease in spectral radius as a function of the number N of nodes and/or the number m of link removals. We also present new bounds for the decrease of the spectral radius of a graph in which m links (or nodes) are removed. In the meanwhile, we find that maximum normalized principal eigenvector component in any graph never exceeds $\sqrt{2}/2$.

Chapter 7 proposes two algorithms, IOPRA and LRA, to generate directed networks with a given directionality ξ . We study the influence of the directionality ξ on the spectral properties of networks. We observe that the spectral radius decreases with the increase of directionality when the degree distribution and the assortativity of the network is preserved. Hereby, we claim that increasing the directionality of directed networks could increase the epidemic threshold to suppress the virus spread. The possible range to increase the epidemic threshold is relatively large in directed binomial networks with a high assortativity and directed power-law networks with a low assortativity. We also find that the influence of each node on the spectral radius is similar in networks with a high directionality. Moreover, we show that the spectral gap and the algebraic connectivity increase with the directionality, implying that an increase of the directionality enhances

the connectivity of the network. We conclude that in directed networks the epidemic threshold is larger and a random walk converges to its steady-state faster than that in undirected networks with the same degree distribution.

8.2 Contributions Summary

The most important contributions of this thesis are highlighted as follows:

- We find that a network with a given degree distribution can indeed be characterized by a small representative set of metrics sufficiently and effectively. Most metric correlations are explained analytically. Moreover, we find that the spectral metrics seem to be essential in network characterizations.
- We study relation between the degree and the principal eigenvector, and find that the difference between the principal eigenvector and the scaled degree eigenvector is the smallest, when $\lambda_1 = N_2/N_1$. What is more, the linear correlation between the degree and the principal eigenvector is strong when the assortativity is large.
- A novel centrality measure, the degree mass, is introduced in this thesis. The correlations between centrality metrics are studied via the Pearson correlation coefficient and the centrality similarity. We find that the betweenness, the closeness and the components of the principal eigenvector are strongly correlated with the degree, the 1st-order degree mass and the 2nd-order degree mass, respectively. Furthermore, the $\rho(x_1, D^{(2)}) > \rho(x_1, D^{(1)}) > \rho(x_1, D)$ is theoretically proven for ER networks.
- Centrality metrics are applied to inflexible contrarian opinion (ICO) model to help one opinion to compete with the other. We learn that the leverage, the betweenness or the degree strategy is more effective in opinion-competition than using other centrality strategies in all types of networks.
- An ε -SIS spreading model is proposed, and taken as a benchmark for the comparison between two mean-field approximations, the NIMFA and the HMF approximation of the SIS model. We find that, overall, the NIMFA is superior to the HMF approximation. An analytic comparison of the epidemic thresholds of the two mean-field approximations is studied.
- The decrease of the spectral radius, an important characterization of network dynamics, by removing links (or nodes) is investigated. The strategy that removes the node u or the link $l = i \sim j$ with highest vector component $(x_1)_u$ or largest product $(x_1)_i(x_1)_j$ of the components of x_1 is superior to other strategies in decreasing λ_1 in most cases.
- A new type of bounds for the spectral radius λ_1 of a graph in which m nodes or links are removed are presented.
- We propose two algorithms, IOPRA and LRA, to generate directed networks with a given directionality ξ , which allow us to study the influence of the directionality ξ on the spectral properties of networks. A universal observation is that, the spectral radius decreases with the directionality when the degree distribution and the assortativity of the network is preserved.

8.3 Directions for Future Work

In this thesis we have thoroughly studied various problems related to characterizing and designing complex networks. Although this work is a tip of the iceberg in complex network field, we believe it will open a door to a wide range of related future works. Here we identify several interesting research areas for future works, which can be direct extensions of this thesis or independent work that may result in another thesis.

1. Chapter 3 shows that a small representative set of metrics can characterize a network with given degree distribution. However, the representative set tends to be larger when more than 9 metrics are considered. So far, we have explored the effect of degree distribution on metric correlations. It is interesting to further examine a large set of metrics and the metric correlation pattern in graphs with a given degree distribution and a given degree correlation (assortativity).

2. Topologies of many complex networks such as the Internet at router level are not known due to their large size or technical challenges, although some properties of their sub-networks can be measured. The dependency of metric correlation pattern on network properties opens up a new direction of reverse-engineering: infer possible universal network properties or the possible suitable network model of a class of graphs from the correlations of the measured network metrics.

3. The epidemic processes studied in Chapters 5-7 are homogeneous. The infection rate on all links between infected and susceptible nodes is the same β , and the curing rate for infected nodes is the same δ . An interesting research direction is to investigate the heterogeneous epidemic processes, where the spreading rates for links are different and the curing rates for infected nodes are different. Understanding how the epidemic threshold alters with the spreading and curing rates would improve the knowledge of epidemic processes.

4. In Chapter 7, the effect of the directionality on the spectral properties is studied in directed networks with the same in- and out- degree distribution. The influence of the difference between in- and out- degree of nodes on spectral properties for directed power-law networks remains an open question.

5. Due to the critical effects of cascading failures, the study of network of networks (also called the interdependent networks) has become a hot topic and caught the attention of the scientific community. To date, most studies focus on the network metrics, the epidemic spreading, and the cascading failures in two-layer interdependent networks. Many properties of multi-layer interdependent networks remain unexplored.

6. The need for exactly modeling real-world networks is particularly intense. There is much to be done in developing more sophisticated models of networks. Better understanding the properties of networks can help to model the real-world networks better. For example, brain networks often are small-world (high clustering; short path lengths) and scale-free, but they have other properties that are not explained by these two models: for example, the modularity and the assortativity. Building up a precise model for a class of real-world networks is in hurry, since the model is the law governing the dynamic processes.

APPENDIX A

Structural Metric Correlations in Network Models

The following plots offer additional material to that of Section 3. The images depict the relation between the structural metrics of network models (Erdős-Rényi random graphs, Barabasi-Albert graphs and Watts-Strogatz Small-world graphs).

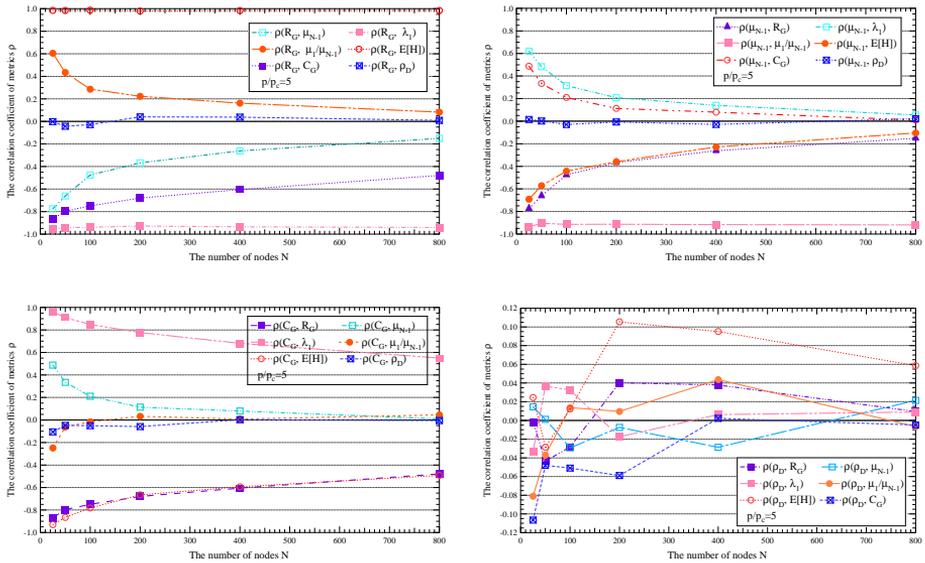


Figure A.1: Metric correlations as a function of N in Erdős-Rényi random graphs ($p = 5pc$).

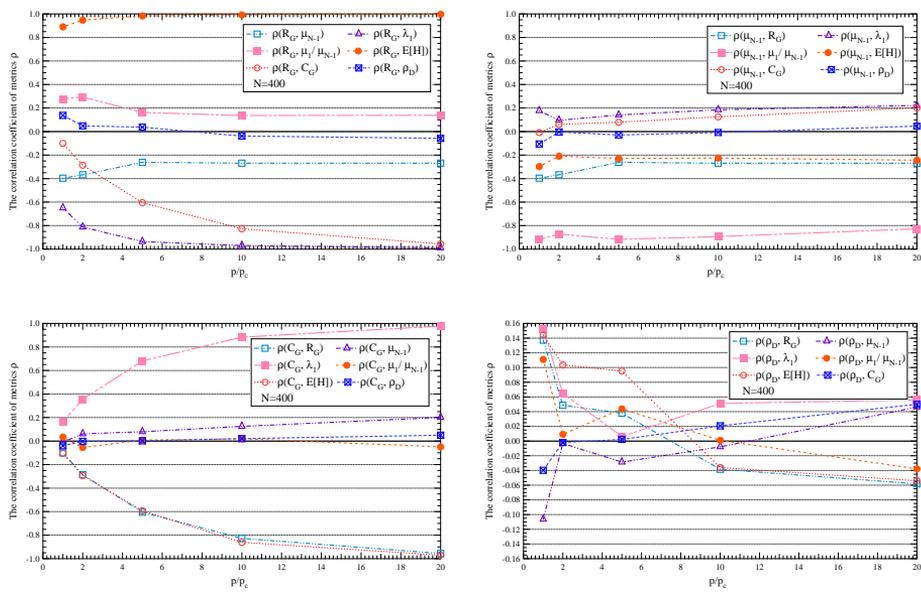


Figure A.2: Metric correlations as a function of p/p_c in Erdős-Rényi random graphs ($N = 400$).

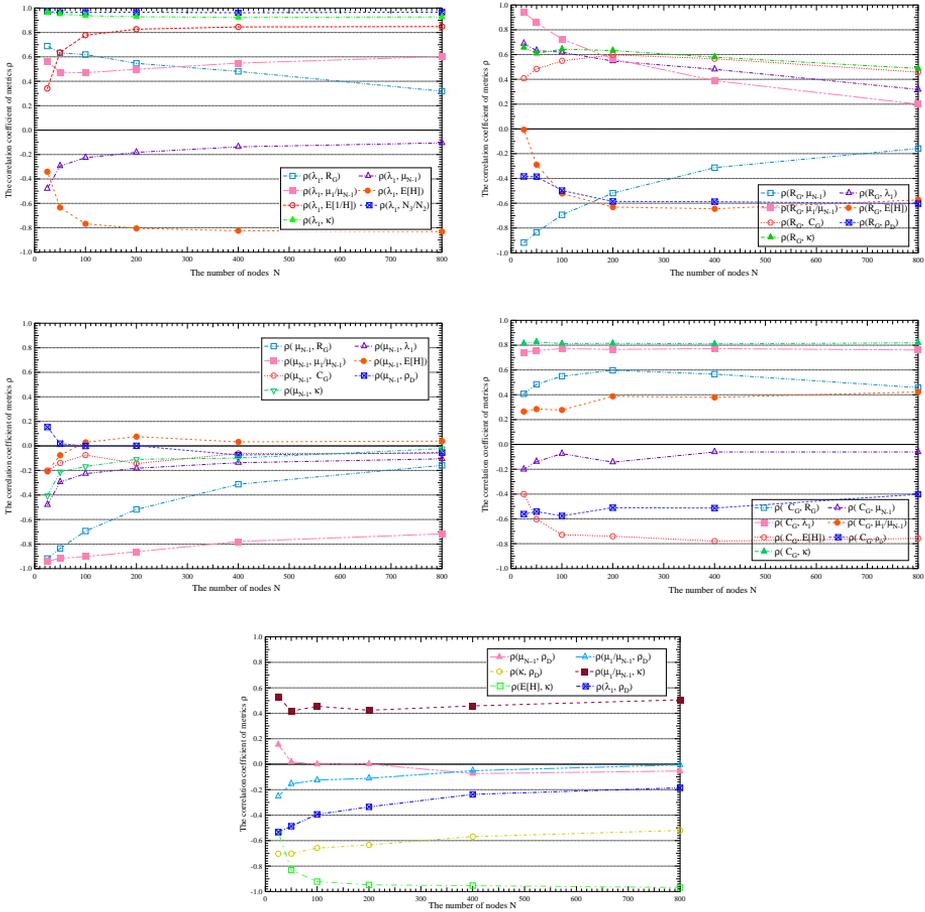


Figure A.3: Metric correlations as a function of N in Barabasi-Albert graphs ($m = 4$).

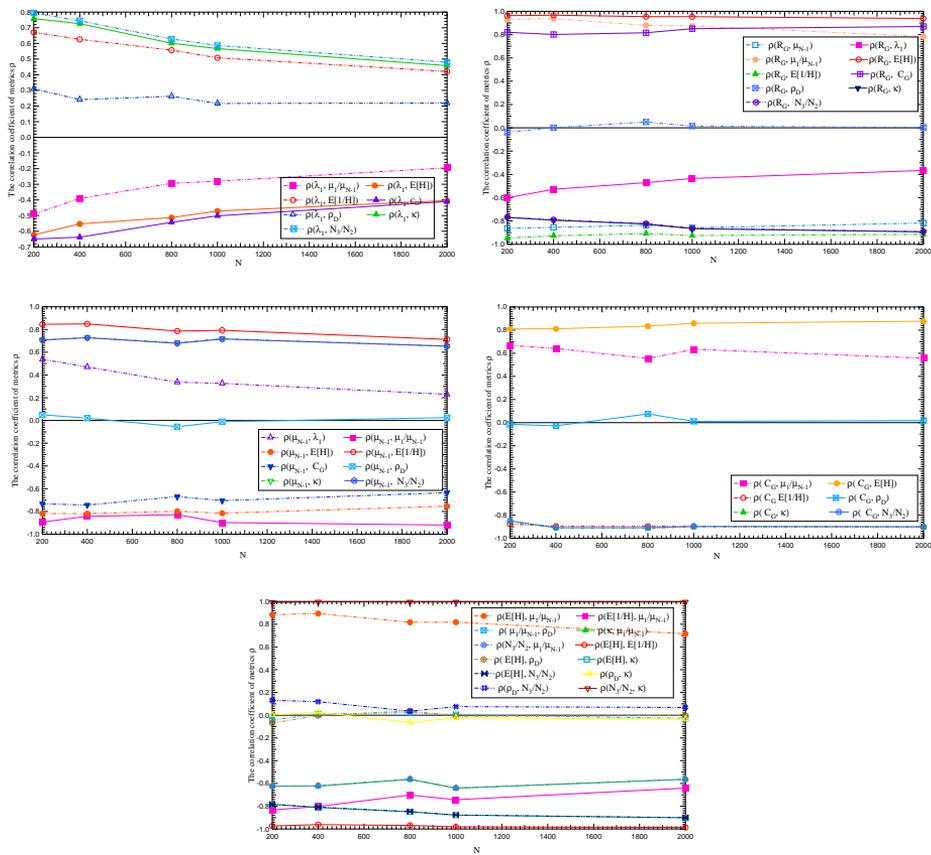


Figure A.4: Metric correlations as a function of N in WS Small-world graphs ($k = 6, p_r = 0.01$).

APPENDIX B

Introduction of Real-world Networks

B.1 Descriptions of the undirected real-world networks studied in Section 4.3

The real-world networks studied in Section 4.3 are described in Table B.1. The real-world networks are indexed in the first column.

B.2 Properties of the undirected real-world networks studied in Section 4.3

The properties of real-world networks are shown in the Table B.2. The definition of these properties has been described in detail in Section 2.1.

Table B.1: *Descriptions of real-world networks.*

Index	Networks	Descriptions
1	American airline	The direct airport-to-airport American mileage a maintained by the U.S. Bureau of Transportation Statistics.
2	American football	This is the network of American football games between Division IA colleges during regular season Fall 2000, as compiled by M. Girvan and M. Newman.
3	ARPANET80	The Advanced Research Projects Agency Network as seen in 1980.
4	Celegensneural	Network representing the neural network of C. Elegans.
5	Dophins	An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand.
6	Dutch soccer	Dutch football players represent the nodes. Two nodes are linked if they played together a match.
7	Gnutella 1	Gnutella snapshots. Four different crawls are available.
8	Gnutella 2	
9	Gnutella 3	
10	Gnutella 4	
11	Karate	Social network of friendships between 35 members of a karate club at a US university in the 1970.
12	LesMis	Coappearance network of characters in the novel Les Miserables.
13	Surfnet	SURFNET topology inferred from the switch interface interconnections.
14	Electric s208	ISCAS89 Sequential Benchmark Circuits. Each node represents a logical operation implemented physically. Links between them relate their inputs/outputs.
15	Electric s420	
16	Electric s838	
17	Epowergrid1	Power-grid infrastructure at three different levels of one city-area in Western Europe.
18	Epowergrid2	
19	Epowergrid3	
20	Erailway1	Railway infrastructure at two levels of one Western-European country
21	Erailway2	
22	WordAdj	Adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens.
23	WordAdjEnglish	Word-adjacency networks of texts in English, French and Japanese separately.
24	WordAdjFranch	
25	WordAdjJapanese	
26	Internet AS (01')	Internet snapshot retrieved from the merge of different data sources (BGP routing tables and updates: Route Views, RIPE, Abilene, CERNET, BGP View).
27	Astro_Ph	Network of coauthorships between scientists posting preprints on the Astrophysics E-Print Archive between Jan 1, 1995 and December 31, 1999.
28	SciMet	Web of Science C. The citation network was created using the Web of Science database SciMet. Networks created with the tool HistCite.
29	HighE-th	High Energy Theory C. Network of coauthorships between scientists posting preprints on the High-Energy Theory E-Print Archive between Jan 1, 1995 and December 31, 1999.
30	CondMat 95-03	Network of coauthorships between scientists posting preprints on the Condensed Matter E-Print Archive. We have two networks corresponding to different periods of time. Periods are Jan 1, 1995 - December 31, 1999 and 2003 respectively.
31	CondMat 95-99	
32	Dutch Roadmap	A graph representing the interconnection between cities in the Netherlands.
33	Network Science C	Coauthorship network of scientists working on network theory and experiment, as compiled by M. Newman in May 2006.
34	Next Generation	A typical Next Generation Transport network.

Table B.2: Properties of real-world networks. The real-world network index is shown in Table B.1. N is the number of nodes, L is the number of links. $E[H]$ is the average shortest path, C_G is the clustering coefficient of networks, ρ_D is the degree correlation coefficient (called the assortativity) of networks. λ_1 is the largest eigenvalue (called spectral radius) of the adjacency matrix of the network. μ_{N-1} is the second smallest Laplace eigenvalue (called spectral radius) of the networks. μ_1/μ_{N-1} is the ratio of the largest eigenvalue μ_1 and the second smallest eigenvalue μ_{N-1} of Laplacian matrix. R_G is the effective graph resistance.

Index	N	L	$E[H]$	C_G	ρ_D	λ_1	μ_{N-1}	μ_1/μ_{N-1}	R_G	$E[D]$	$\sqrt{\text{Var}[D]}$	H_{max}
1	2179	31326	3.0262	0.4849	-0.0409	144.6112	0.2082	2.0675e ³	1.6072e ⁴	28.7526	56.6782	8
2	115	613	2.5082	0.4032	0.1624	10.7806	1.4590	1.5086e ³	10.6609	0.8835	0.8835	4
3	71	86	6.4849	0.0141	-0.2613	2.7648	0.0374	170.2063	7.0158e ³	2.42254	0.7442	17
4	297	2148	2.4553	0.2924	-0.1632	24.3655	0.8485	159.1562	1.3710e ⁴	14.4646	12.9443	5
5	62	159	3.3570	0.2590	-0.0436	7.1936	0.1730	78.7034	1.8643e ³	5.1290	2.9319	8
6	685	10310	4.4583	0.7506	-0.0634	50.8428	0.1613	372.0373	3.1157e ⁴	30.1022	21.1957	11
7	737	803	9.1351	0.0063	-0.1934	4.8913	0.0073	2.6292e ³	1.4181e ⁶	2.1791	2.0069	24
8	1568	1906	6.1037	0.0192	-0.0946	13.7828	0.0167	1.1205e ⁴	4.0212e ⁴	2.4311	5.5778	21
9	435	459	6.7085	0.0145	-0.3301	8.2281	0.0110	5.9278e ³	4.2538e ⁵	2.1103	5.1534	20
10	653	738	5.4513	0.0232	-0.2459	12.1145	0.0231	6.2319e ³	6.6603e ⁵	2.2603	7.0228	15
11	35	134	1.9126	0.3908	-0.5036	9.6253	1.7264	12.6030	221.6283	7.6571	4.7265	3
12	77	254	2.6411	0.5731	-0.1652	12.0058	0.2050	180.9490	3.0166e ³	6.5974	6.0006	5
13	65	111	4.1236	0.0359	0.2288	5.0523	0.1137	92.7068	3.2979e ³	3.4154	1.9046	10
14	122	189	4.9278	0.0591	-0.0020	4.1036	0.0836	135.2786	1.3082e ⁴	3.0984	1.4395	11
15	252	399	5.8064	0.0651	-0.0059	4.3600	0.0512	297.3970	5.8313e ⁴	3.1667	1.5340	13
16	512	819	6.8585	0.0547	-0.0300	5.0097	0.0285	809.9553	2.5149e ⁵	3.1992	1.6296	15
17	3419	3953	21.1147	0.0120	-0.1283	5.1781	< e ⁻⁵	> e ¹⁵	4.8953e ⁷	2.31237	1.84246	51
18	1205	1384	12.3547	0.0171	0.1082	4.8994	0.0022	9.1191e ³	4.3901e ⁶	2.2971	1.3609	31
19	395	441	13.6088	0.0201	-0.0235	4.4854	0.0020	8.8844e ³	7.2535e ⁵	2.2329	1.2834	42
20	8710	11332	79.0448	0.0212	-0.0219	2.9865	< e ⁻⁵	> e ¹⁵	7.2107e ⁸	2.6021	0.7696	213
21	689	778	34.1261	0.0731	0.0980	3.6926	7.7321e ⁻³	1.0526e ⁴	3.9229e ⁶	2.2583	0.7658	84
22	112	425	2.5356	0.1728	-0.1293	13.1502	0.6950	72.0767	3.7941e ³	7.5893	6.8512	5
23	7377	44205	2.7780	0.4085	-0.2366	109.4416	< e ⁻⁵	9.1266e ¹⁵	2.2149e ⁷	11.9846	60.8260	8
24	8308	23832	3.2189	0.2138	-0.2330	60.6735	0.1197	1.5810e ⁴	3.9917e ⁷	5.7371	34.8979	9
25	2698	7995	3.0771	0.2196	-0.2590	42.9980	< e ⁻⁵	5.8851e ¹⁵	4.3489e ⁶	5.9266	24.6695	8
26	12254	25319	3.6214	0.2992	-0.1903	61.1066	< e ⁻⁵	4.8974e ¹⁵	1.0349e ⁸	4.1324	33.5463	11
27	14845	119652	4.7980	0.6696	0.2277	73.8868	0.0302	1.1966e ⁴	7.2012e ⁷	16.1202	21.7466	14
28	2678	10368	4.1797	0.1736	-0.0352	20.4290	0.0853	1.9365e ⁶	2.9549e ⁶	7.7431	9.2480	12
29	5835	13815	7.0264	0.5062	0.1852	18.0442	0.0214	2.3870e ³	2.8800e ⁷	4.7352	4.5571	19
30	27519	116181	5.7667	0.6546	0.1657	40.3097	0.0276	7.3675e ³	3.3638e ⁸	8.4437	10.8110	16
31	13861	44619	6.6278	0.6514	0.1571	24.9822	0.0292	3.6992e ³	1.1613e ⁸	6.4381	6.7598	18
32	29663	34982	148.7102	0.0443	0.2462	3.4567	< e ⁻⁵	> e ¹⁵	1.5472e ¹⁰	2.3586	0.6823	531
33	379	914	6.0419	0.7412	-0.0817	10.3755	0.0152	2.3053e ³	1.4526e ⁵	4.8232	3.9272	17
34	29902	32707	7109.8681	0.0306	-0.0355	49.5455	< e ⁻⁵	> e ¹⁵	2.1188e ¹²	2.1876	9.7574	14253

B.3 Descriptions of the directed real-world networks studied in Chapter 7

- Enron:

This data set was made public by the Federal Energy Regulatory Commission during its investigations: it is a partially anonymised corpus of e-mail messages exchanged by some Enron employees (mostly part of the senior management). This data set is a directed graph, whose nodes represent people and with an arc from x to y whenever y was the recipient of (at least) a message sent by x .

- Ljournal-2008:

LiveJournal is a virtual-community social site started in 1999: nodes are users and there is an arc from x to y if x registered y among his friends. It is not necessary to ask y permission, so the graph is directed. This graph is the snapshot used by Chierichetti, Flavio and *et al.* in “On compressing social networks.” Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009, and was kindly provided by the authors.

- Twitter-2010

Twitter is a website, owned and operated by Twitter Inc., which offers a social networking and microblogging service, enabling its users to send and read messages called tweets. Tweets are text-based posts of up to 140 characters displayed on the user’s profile page. This is a crawl presented by Kwak, Haewoon and *et al.* in “What is Twitter, a social network or a news media?”, Proceedings of the 19th international conference on World wide web. ACM, 2010. Nodes are users and there is an arc from x to y if y is a follower of x . In other words, arcs follow the direction of tweet transmission.

- Word Association-2011

The Free Word Association Norms Network is a directed graph describing the results of an experiment of free word association performed by more than 6000 participants in the United States: its nodes correspond to words and arcs represent a cue-target pair (the arc $x \rightarrow y$ means that the word y was output by some of the participants based on the stimulus x).

- WWW networks

The networks, “cnr-2000”, “in-2004”, “eu-2005”, “uk-2007-05@100000” and “uk-2007-05@1000000” are small WWW networks that were crawled from the Internet. The “cnr-2000” is crawled from the Italian CNR domain. A small crawl of the .in domain performed for the Nagaoka University of Technology is in data “in-2004”. The “eu-2005” is a small crawl of the .eu domain. This network “uk-2007-05@100000” and “uk-2007-05@1000000” have been artificially generated by combining twelve monthly snapshot of the .uk domain and collected for the DELIS project.

APPENDIX C

Pearson Correlation Coefficient Between Centrality Metrics

This appendix offers additional material to that of Section 4. The sequence of figures shows the Pearson correlation coefficient between any two centrality metrics in network models (Erdős-Rényi random networks and scale-free networks). The effect of the size of SF and the link density of ER on the correlation coefficients is illustrated in the figures. The following table displays the Pearson correlation coefficient between any two centrality metrics in 34 real-world networks, which are described in Appendix B.

The correlation indexes mentioned in the following images and tables are the indexes for pairs of centrality metrics: 1. (B_n, C_n) ; 2. (B_n, D) ; 3. (B_n, x_1) ; 4. (B_n, K_s) ; 5. (B_n, L_n) ; 6. $(B_n, D^{(1)})$; 7. $(B_n, D^{(2)})$; 8. (C_n, D) ; 9. (C_n, x_1) ; 10. (C_n, K_s) ; 11. (C_n, L_n) ; 12. $(C_n, D^{(1)})$; 13. $(C_n, D^{(2)})$; 14. (D, x_1) ; 15. (D, K_s) ; 16. (D, L_n) ; 17. $(D, D^{(1)})$; 18. $(D, D^{(2)})$; 19. (x_1, K_s) ; 20. (x_1, L_n) ; 21. $(x_1, D^{(1)})$; 22. $(x_1, D^{(2)})$; 23. (K_s, L_n) ; 24. $(K_s, D^{(1)})$; 25. $(K_s, D^{(2)})$; 26. $(L_n, D^{(1)})$; 27. $(L_n, D^{(2)})$; 28. $(D^{(1)}, D^{(2)})$.

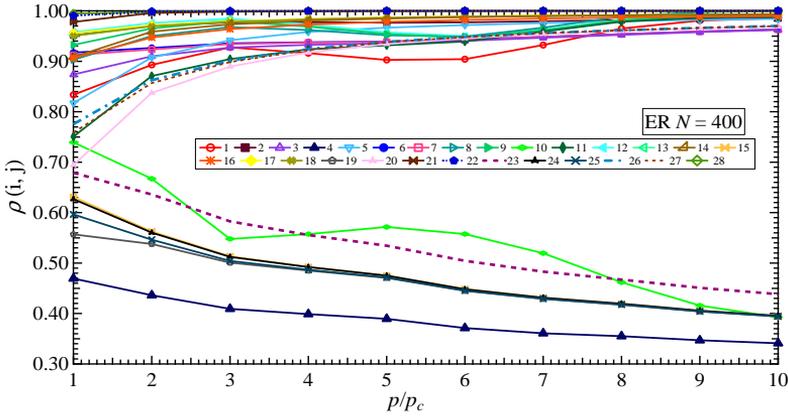


Figure C.1: Pearson correlation coefficient between any two centrality metrics as a function of p/p_c , in ER networks ($N = 400$). The number in the annotation is the correlation index.

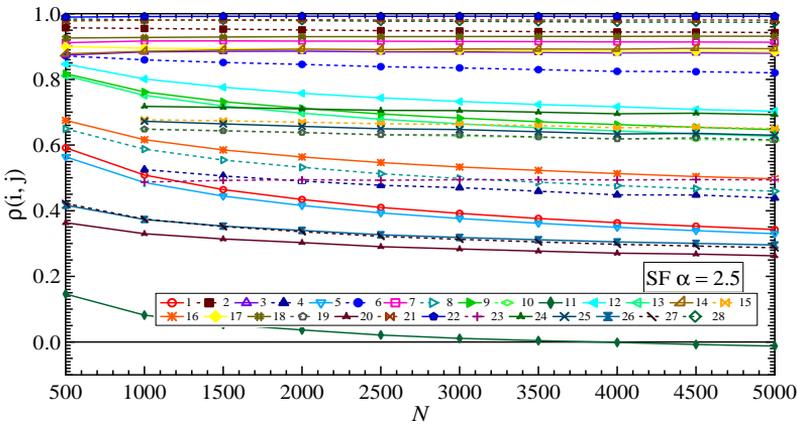


Figure C.2: Pearson correlation coefficient between any two centrality metrics as a function of the size N of networks, in scale-free networks ($\alpha = 2.5$). The number in the annotation is the correlation index.

Table C.1: *Pearson correlation coefficients among the centrality metrics in the real-world networks. The horizontal index is the correlation index, and the vertical index is the real-world network index.*

Index	1	2	3	4	5	6	7	8	9
1	0.3667	0.5690	0.4119	0.3377	0.4027	0.4314	0.4224	0.7580	0.7684
2	0.8167	0.2813	0.1450	0.0871	0.3212	0.2230	0.2075	0.2913	0.2462
3	0.7129	0.7235	0.5358	0.3496	0.55585	0.7660	0.7593	0.4308	0.6851
4	0.4271	0.7805	0.5206	0.1822	0.4212	0.5388	0.6044	0.6997	0.7827
5	0.6657	0.5902	0.2835	0.4703	0.5639	0.5131	0.4850	0.7127	0.6979
6	0.3303	0.4909	0.0857	0.1523	0.4170	0.3807	0.3113	0.2701	-0.1604
7	0.4456	0.7292	0.4780	0.5182	0.4556	0.7575	0.7882	0.3973	0.5241
8	0.2196	0.9691	0.7006	0.2677	0.2679	0.3858	0.9416	0.2225	0.5469
9	0.2475	0.8839	0.4926	0.4667	0.4356	0.3533	0.8283	0.1763	0.5112
10	0.2338	0.9603	0.5848	0.3296	0.3880	0.2640	0.8839	0.1774	0.5733
11	0.8699	0.9651	0.8757	0.3782	0.8707	0.7999	0.9166	0.8853	0.9599
12	0.6287	0.7468	0.4231	0.2388	0.5317	0.5534	0.5468	0.7997	0.6812
13	0.7136	0.8743	0.7365	0.6345	0.6985	0.7999	0.7816	0.8290	0.9286
14	0.6408	0.7475	0.5595	0.2147	0.5551	0.7357	0.7227	0.6127	0.7987
15	0.5956	0.6933	0.5514	0.1583	0.4508	0.7084	0.7203	0.5541	0.7178
16	0.5323	0.7044	0.5410	0.1314	0.3913	0.6971	0.7661	0.4623	0.5633
17	0.2349	0.3843	0.1180	0.1189	0.1889	0.4101	0.4082	0.1082	0.0607
18	0.3210	0.7005	0.5517	0.0560	0.2686	0.6772	0.7144	0.2946	0.4627
19	0.3001	0.7081	0.4775	0.1060	0.2945	0.6371	0.6825	0.2395	0.4925
20	0.2664	0.1565	-0.0442	0.1979	0.1112	0.1805	0.1876	0.1477	0.0209
21	0.5022	0.3274	0.0364	0.3836	0.2548	0.2790	0.2540	0.2428	0.1141
22	0.6559	0.9150	0.8226	0.3517	0.6586	0.7891	0.8444	0.8410	0.9245
23	0.1880	0.9225	0.6525	0.2068	0.2642	0.4157	0.7765	0.3535	0.6528
24	0.1874	0.9714	0.8047	0.2729	0.2636	0.4403	0.9385	0.2625	0.6215
25	0.2747	0.9660	0.7859	0.3249	0.3584	0.5266	0.8972	0.3868	0.6880
26	0.1382	0.9826	0.7994	0.3292	0.2290	0.3441	0.9582	0.1631	0.5776
27	0.3764	0.6787	0.4353	0.2869	0.4631	0.5670	0.5270	0.6109	0.4220
28	0.4068	0.8185	0.6959	0.3147	0.4401	0.7143	0.7605	0.6741	0.7030
29	0.4526	0.7798	-0.0109	0.3574	0.5079	0.6700	0.5803	0.5774	0.0119
30	0.3801	0.7534	0.3753	0.3152	0.4488	0.5933	0.5173	0.5989	0.3906
31	0.4002	0.7225	0.2781	0.2607	0.4581	0.5718	0.4816	0.5616	0.3248
32	0.2214	0.1741	-0.0037	0.1619	0.1117	0.1719	0.1608	0.1450	-0.0221
33	0.4302	0.6883	0.1884	0.1917	0.4707	0.5630	0.4997	0.3468	0.2593
34	-0.1342	-0.0436	-0.6295	-0.9718	0.9538	-0.9051	-0.1342	0.0313	0.2446

Index	10	11	12	13	14	15	16	17	18	19
1	0.8174	0.5944	0.7903	0.7712	0.9592	0.8730	0.7259	0.9657	0.9643	0.9254
2	0.1742	0.2704	0.2826	0.2839	0.7501	0.3881	0.9181	0.9619	0.9314	0.2456
3	0.3807	0.2524	0.5598	0.5870	0.4650	0.5127	0.8914	0.9020	0.9079	0.1326
4	0.6861	0.5776	0.8680	0.7951	0.7810	0.5434	0.7886	0.8830	0.9311	0.5572
5	0.7498	0.6094	0.7475	0.7422	0.7196	0.8303	0.9050	0.9574	0.9417	0.5388
6	0.0680	0.2221	0.2381	0.1801	0.6237	0.7300	0.8963	0.9393	0.8801	0.7983
7	0.5073	0.2052	0.6184	0.6248	0.4660	0.5933	0.8117	0.8217	0.8573	0.3912
8	0.4015	0.0017	0.7515	0.3210	0.6523	0.3463	0.3888	0.3594	0.9132	0.1840
9	0.2377	-0.3534	0.8326	0.3544	0.5811	0.3316	0.4651	0.3050	0.9493	0.2032
10	0.2234	-0.2234	0.8594	0.2967	0.6366	0.2492	0.3751	0.2256	0.9481	0.0868
11	0.5492	0.7227	0.9606	0.9463	0.9392	0.5331	0.9390	0.8718	0.9714	0.6221
12	0.5622	0.6340	0.8375	0.7931	0.8467	0.7969	0.8474	0.9455	0.9380	0.8100
13	0.7311	0.3466	0.9330	0.9363	0.9046	0.8289	0.7598	0.9486	0.9391	0.8425
14	0.5670	0.3265	0.7388	0.7574	0.6757	0.4296	0.8184	0.9260	0.9225	0.3108
15	0.5257	0.2675	0.6964	0.7100	0.6147	0.3995	0.7980	0.9078	0.9200	0.2464
16	0.4949	0.1937	0.6534	0.6411	0.4120	0.3738	0.7690	0.8670	0.9055	0.1143
17	-0.0402	-0.0122	0.1651	0.1653	0.2143	0.4102	0.6878	0.7733	0.8456	0.0447
18	0.1582	0.1027	0.4752	0.4902	0.5040	0.1904	0.5901	0.8725	0.8851	0.0638
19	0.2490	0.2137	0.5599	0.5316	0.5183	0.2287	0.5911	0.7611	0.8338	0.0327
20	0.1649	0.0836	0.1829	0.2016	0.1031	0.7905	0.9247	0.9522	0.9241	0.1132
21	0.4880	0.0325	0.3314	0.3382	0.2678	0.4149	0.7508	0.8884	0.8524	0.0835
22	0.8194	0.7371	0.9451	0.9123	0.9575	0.6433	0.8327	0.9390	0.9707	0.7010
23	0.7195	0.3891	0.8312	0.5353	0.8704	0.4649	0.4862	0.6580	0.9504	0.7992
24	0.6355	0.0669	0.8167	0.4111	0.8733	0.4146	0.3627	0.5403	0.9779	0.6980
25	0.6814	0.2080	0.8410	0.5506	0.8911	0.5155	0.5048	0.6631	0.9694	0.7628
26	0.4291	-0.0707	0.7971	0.2788	0.8253	0.3935	0.2696	0.3771	0.9754	0.5413
27	0.5427	0.2819	0.5861	0.5264	0.7188	0.8070	0.5920	0.9352	0.8728	0.5695
28	0.8188	0.5093	0.7923	0.7456	0.8345	0.6962	0.7237	0.9204	0.9236	0.6212
29	0.4884	0.2103	0.6517	0.6022	0.1789	0.7311	0.7080	0.9080	0.8292	0.5171
30	0.6341	0.2404	0.6153	0.5392	0.6346	0.7339	0.6197	0.9035	0.8259	0.5001
31	0.5157	0.2077	0.6067	0.5300	0.5304	0.7166	0.6631	0.8941	0.8021	0.4229
32	0.1465	-0.0170	0.2033	0.2220	0.0364	0.5291	0.7674	0.9271	0.8880	0.0101
33	0.0926	0.0970	0.4562	0.4120	0.4748	0.6803	0.7723	0.8795	0.8415	0.4195
34	0.3609	-0.3531	0.3378	0.0649	0.7297	0.0866	0.0487	0.1570	0.9858	0.6768

Index	20	21	22	23	24	25	26	27	28
1	0.6327	0.9978	0.9998	0.7122	0.9389	0.9245	0.6604	0.6405	0.9984
2	0.4881	0.8660	0.9134	0.4481	0.3467	0.3274	0.7771	0.7189	0.9929
3	0.1934	0.6460	0.7101	0.5798	0.4773	0.4407	0.6485	0.6530	0.9811
4	0.6130	0.9783	0.9885	0.7710	0.6277	0.5737	0.6605	0.6789	0.9813
5	0.4991	0.8285	0.8842	0.8506	0.8171	0.7668	0.7887	0.7535	0.9913
6	0.3684	0.8132	0.8867	0.6089	0.8563	0.8700	0.7478	0.6517	0.9864
7	0.2262	0.6736	0.7412	0.4906	0.6480	0.5920	0.5024	0.4922	0.9475
8	0	0.8135	0.8463	0.5030	0.3187	0.2061	-0.0050	0.1176	0.4936
9	-0.1161	0.7007	0.7440	0.3782	0.2365	0.2762	-0.3636	0.2134	0.4889
10	-0.1437	0.7414	0.8018	0.5184	0.1290	0.1398	-0.3598	0.1438	0.3751
11	0.8128	0.9837	0.9930	0.5722	0.6484	0.5928	0.7290	0.8623	0.9568
12	0.6520	0.9427	0.9691	0.7984	0.8524	0.8447	0.7713	0.7455	0.9924
13	0.4673	0.9841	0.9927	0.6005	0.8604	0.8512	0.5510	0.5248	0.9969
14	0.3310	0.8087	0.8589	0.2983	0.4885	0.4576	0.6007	0.5809	0.9839
15	0.2497	0.7503	0.8010	0.2684	0.4523	0.4137	0.5520	0.5475	0.9788
16	0.0562	0.5789	0.6656	0.2530	0.4262	0.3673	0.4862	0.4847	0.9533
17	0.0545	0.3371	0.3805	0.7626	0.1513	0.1393	0.2283	0.2760	0.9458
18	0.0501	0.6365	0.6794	0.3420	0.1429	0.1199	0.2204	0.2123	0.9812
19	0.0748	0.7433	0.7697	0.3351	0.1335	0.1077	0.0448	0.1010	0.9619
20	0.0564	0.1303	0.1454	0.6233	0.8541	0.8624	0.7665	0.7184	0.9907
21	0.0347	0.4062	0.4780	0.2918	0.4205	0.3829	0.4013	0.3398	0.9842
22	0.7490	0.9949	0.9983	0.8031	0.7300	0.6910	0.7541	0.7622	0.9888
23	0.6646	0.9320	0.9790	0.7406	0.8912	0.6890	0.6611	0.6156	0.8432
24	0.3912	0.8774	0.9476	0.5641	0.7939	0.5488	0.3408	0.3734	0.6794
25	0.5507	0.9242	0.9721	0.6990	0.8180	0.6646	0.4857	0.5386	0.8112
26	0.1486	0.8169	0.8977	0.4876	0.5646	0.4417	0.0699	0.1943	0.4845
27	0.2248	0.8789	0.9367	0.4761	0.7840	0.7124	0.3996	0.3245	0.9845
28	0.4680	0.9417	0.9682	0.7181	0.7457	0.6886	0.5866	0.5501	0.9877
29	0.0427	0.2885	0.3822	0.5164	0.7657	0.7361	0.4493	0.3477	0.9771
30	0.1765	0.8431	0.9205	0.5016	0.7344	0.6617	0.3726	0.2850	0.9795
31	0.1358	0.7641	0.8725	0.4877	0.7372	0.6597	0.3945	0.2903	0.9731
32	0.0063	0.0524	0.0629	0.3943	0.5167	0.4740	0.4892	0.4156	0.9878
33	0.1267	0.7062	0.8105	0.5701	0.7390	0.6966	0.5089	0.4324	0.9766
34	-0.5920	0.7797	0.8022	-0.9766	0.9347	0.1797	-0.9156	-0.0611	0.2549

APPENDIX D

Algorithms for Generating Directed Networks with Given Directionality

This appendix provides two algorithms, In-degree and Out-degree Preserving Rewiring Algorithm (IOPRA) and Link resetting algorithm (LRA), which both can be applied to any network to generate a directed network with a given directionality.

Algorithm 1 *IOPRA*(G, ξ)

```
1: Create a bidirectional network  $G(N, L)$ ;  
2: Save network  $G(N, L)$  as  $G_s$  and calculate the directionality  $\xi_s$  of network  $G_s$ ;  
3: while  $|\xi_s - \xi| > 10^{-5}$  do  
4:   Randomly select two unidirectional links  $i \rightarrow j$  and  $k \rightarrow l$  associated with the four nodes  $i, j, k, l$ ;  
5:   Rewire the link pair  $i \rightarrow j$  and  $k \rightarrow l$  into  $i \rightarrow l$  and  $k \rightarrow j$ . The new network  $G_n$  is obtained;  
6:   Calculate the directionality  $\xi_n$  of the network  $G_n$ ;  
7:   if  $|\xi_s - \xi| > |\xi_n - \xi|$  then  
8:      $G_s \leftarrow G_n$ ;  
9:      $\xi_s \leftarrow \xi_n$ ;  
10:  else  
11:    Give up this rewired node pair;  
12:  end if  
13: end while  
14: return  $G_s$ 
```

Algorithm 2 *LRA*(G, ξ)

```
1: Create a bidirectional network  $G(N, L)$ ;  
2: Randomly choose  $\xi$  percentage of bidirectional link pairs;  
3: Randomly choose one unidirectional link from each link pair;  
4: Randomly reset the chosen unidirectional links to the locations without any link;  
5: Save the new network as  $G_s$ ;  
6: return  $G_s$ 
```

APPENDIX E

Eigenvalues of the Directed Networks

The spectral radius λ_1 and the spectral gap ($\lambda_1 - \lambda_2$) are considered as important metrics for the percolation processes on networks. Here we also present all eigenvalues in directed networks in a Image-Real figure. The eigenvalues are calculated on 10^3 simulation realizations. The changes of the eigenvalues λ_i with the directionality ξ from 0 to 1 with step 0.1 in directed binomial networks ($N = 10$, $p = 0.25$) are shown on Figure E.1. Surprisingly, the real part of all eigenvalues tends to 0, when the directionality ξ increases.

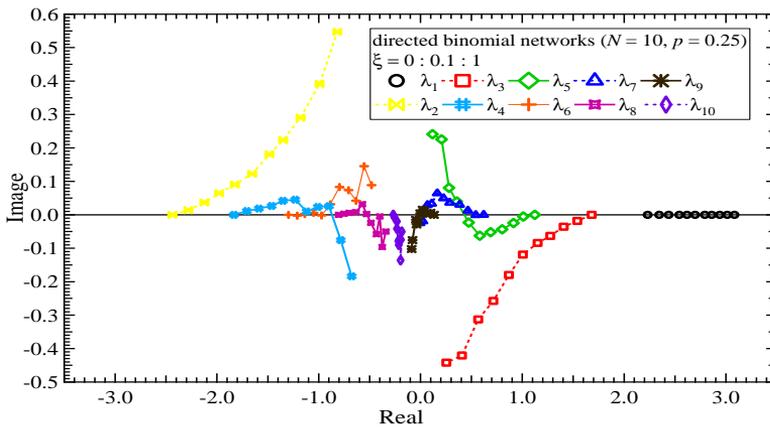


Figure E.1: Change of the eigenvalues with the directionality. When the directionality increases, the real parts of the eigenvalues tend to 0.

List of Abbreviations

BA	Barabási-Albert graph type
CDF	Cumulative Distribution Function
ER	Erdős-Rényi graph type
EEG	Electroencephalograph
HMF	Heterogeneous Mean-Field approximation
ICO	Inflexible Contrarian Opinion model
IOPRA	In-degree and Out-degree Preserving Rewiring Algorithm
IQ	Intelligence Quotient
LRA	Link Resetting Algorithm
LSRM	Link Spectral Radius Minimization
MEG	Magnetoencephalography
NCO	Non-Consensus Opinion model
NIMFA	N-Intertwined Mean-Field Approximation
NP	Non-deterministic Polynomial-time
NSRM	Node Spectral Radius Minimization
SF	Scale-Free graph type
SIR	Susceptible-Infected-Recovered epidemic model
SIS	Susceptible-Infected-Susceptible epidemic model
SL	Synchronization Likelihood
VIP	Very Important Persons
WS	Watts-Strogatz graph type
WWW	World Wide Web

List of Symbols

A	$\begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix}$	adjacency matrix of a graph with N nodes
$\text{diag}(a_k)$	$\text{diag}(a_1, a_2, \dots, a_n)$	a diagonal matrix with diagonal elements listed, while all off-diagonal elements are zero
A^T		transpose of a matrix, the rows of A^T are the columns of A
B		incidence matrix of a graph
B_n		betweenness
C_G		average clustering coefficient
C_i		closeness of node i
D		degree, or diameter
d_i		degree of node i
d_{max} (or d_{min})		maximum (or minimum) degree in a graph
d_{out} (or d_{in})		in-degree (or out-degree) in a directed network
$E[X]$		expectation of the random variable X
G		$= (\mathcal{N}, \mathcal{L})$: a graph
$G_m(\mathcal{L})$ (or $G_m(\mathcal{N})$)		resulting graph after the removal of m links (or nodes) from G
\blacktriangle_G		the number of triangles of a graph
H		hopcount
H_{ij}		number of hops in the shortest path between nodes i and j
H_{max}		maximal hopcount among all node pairs
I		$= \text{diag}(1, 1, \dots, 1)$: identity matrix
K_N		the complete graph with N nodes
$K_{n,m}$		the complete bi-partite graph with $N = n + m$ nodes
k_s		k-shell index
\mathcal{L}		set of links of a graph
\mathcal{L}_m		set of links that removed from a graph
L		$= \mathcal{L} $: number of links in a graph
L_n		leverage centrality

$l(G)$	the line graph of a graph G
$M_{A,B}$	centrality similarity
\mathcal{N}	set of nodes of a graph
$\mathcal{N}^{(k)}$	set of degree k nodes
\mathcal{N}_n	set of nodes that removed from a graph
N	$= \mathcal{N} $: number of nodes in a graph
N_k	$= u^T A^k u$: total number of walks with length k
N_Λ	$= \sum_{i=1}^N \binom{d_i}{2}$: the number of connected triples
P	transition probability matrix (Markov process)
$\text{Prob}[X]$	probability of event X
p	link density
p_c	disconnectivity threshold of Erdős-Rényi random graphs
p_r	rewiring probability of Watts-Strogatz small-world graphs
Q	$= BB^T$ laplacian matrix of a graph
$Q_i(t)$	infinitesimal generator of Markov chain
R_G	effective graph resistance
r	degree of nodes in regular graphs
S_i	disjoint set i of a bipartite graph or the size of the i -th largest cluster
u	all-one vector
$\text{Var}[X]$	$= \sigma_X^2$: variance of the random variable X
$v_i(t)$	probability of node i to be infected at time t
x_i	i -th largest eigenvector (of A or Q , depending on the context)
α	exponent of a power-law degree distribution
β	infection rate over a link in virus spreading
Δ	$\text{diag}(d_1, d_2, \dots, d_N)$: diagonal matrix of the nodal degrees
δ	curing rate for an infected node in virus spreading
ε	spontaneous infection rate for each node in ε -SIS virus spreading
κ	degree diversity
λ_i	i -th largest eigenvalue of the adjacency matrix
λ_1	$= \max_{x^2=1} x^T Q x$, spectral radius
μ_i	i -th largest eigenvalue of the Laplacian matrix
μ_{N-1}	$= \min_{x^2=1, x^T x=0} x^T Q x$, algebraic connectivity
ρ	fraction of infected nodes in a network
ρ_D	assortativity
$\rho(i, j)$	Pearson correlation coefficient
$\rho_k(t)$	probability that a node with k links is infected
$\sigma_{sd}(i)$	number of shortest paths passing through the node i , from node s to node d
σ_{sd}	total number of shortest paths from node s to node d
τ	effective spreading rate
τ_c	epidemic threshold of SIS model
$\tau_c^{(1)}$	epidemic threshold of N-intertwined approximation
τ_c^{HMF}	epidemic threshold of HMF approximation
$\Theta(\rho(t))$	probability that any given link points to an infected node
$\varphi_X(z)$	probability generating function of X
ξ	directionality in a directed network
$1_{\{x\}}$	indicator function: $1_{\{x\}} = 1$ if the event or condition $\{x\}$ is true, else $1_{\{x\}} = 0$

BIBLIOGRAPHY

- [1] P. Erdős and A. Rényi, “On random graphs, i.,” *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.
- [2] P. Erdős and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
- [3] J. Travers and S. Milgram, “An experimental study of the small world problem,” *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.
- [4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the internet topology,” in *ACM SIGCOMM Computer Communication Review*, vol. 29(4), pp. 251–262, ACM, 1999.
- [5] L. A. Adamic and B. A. Huberman, “Power-law distribution of the world wide web,” *Science*, vol. 287, no. 5461, pp. 2115–2115, 2000.
- [6] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [7] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [8] A.-L. Barabasi, *Linked: How everything is connected to everything else and what it means*. Plume Editors, 2002.
- [9] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, Oxford, U.K., 2003.
- [10] R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, Cambridge, U.K., 2007.
- [11] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, U.K., 2008.

- [12] M. E. J. Newman, *Networks: an introduction*. Oxford University Press, Oxford, U.K., 2009.
- [13] R. Cohen and S. Havlin, *Complex networks: structure, robustness and function*. Cambridge University Press, Cambridge, U.K., 2010.
- [14] P. Van Mieghem, *Graph spectra for complex networks*. Cambridge University Press, Cambridge, U.K., 2011.
- [15] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.
- [16] P. Van Mieghem, *Performance analysis of communications networks and systems*. Cambridge University Press, Cambridge, U.K., 2006.
- [17] M. E. J. Newman, “Assortative mixing in networks,” *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.
- [18] M. E. J. Newman, “Mixing patterns in networks,” *Physical Review E*, vol. 67, no. 2, p. 026126, 2003.
- [19] W. Ellens, F. Spijksma, P. Van Mieghem, A. Jamakovic, and R. Kooij, “Effective graph resistance,” *Linear algebra and its applications*, vol. 435, no. 10, pp. 2491–2506, 2011.
- [20] V. Batagelj and M. Zaversnik, “An $o(m)$ algorithm for cores decomposition of networks,” *arXiv preprint cs/0310049*, 2003.
- [21] 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*, vol. 6, no. 11, pp. 888–893, 2010.
- [22] K. E. Joyce, P. J. Laurienti, J. H. Burdette, and S. Hayasaka, “A new measure of centrality for brain networks,” *PLoS One*, vol. 5, no. 8, p. e12200, 2010.
- [23] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang, “Correlation between centrality metrics and their application to the opinion model,” *arXiv preprint: 1409.6033*, 2014.
- [24] F.-G. Juan, S. Krzysztow, J. J. Ramasco, M. S. Miguel, and V. M. Eguíluz, “Is the voter model a model for voters?,” *Physical review letters*, vol. 112, no. 15, p. 158701, 2014.
- [25] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, and C. J. Stam, “The application of graph theoretical analysis to complex networks in the brain,” *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2317–2331, 2007.
- [26] M. Rubinov, S. A. Knock, C. J. Stam, S. Micheloyannis, A. W. Harris, L. M. Williams, and M. Breakspear, “Small-world properties of nonlinear brain activity in schizophrenia,” *Human brain mapping*, vol. 30, no. 2, pp. 403–416, 2009.
- [27] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

- [28] C. J. Stam and J. C. Reijneveld, “Graph theoretical analysis of complex networks in the brain,” *Nonlinear biomedical physics*, vol. 1, no. 1, p. 3, 2007.
- [29] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.
- [30] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, “Epidemic spreading in real networks: An eigenvalue viewpoint,” in *Proceedings of IEEE Symposium on Reliable Distributed Systems, Los Alamitos, CA, USA: IEEE Computer Society*, pp. 25–34, IEEE, 2003.
- [31] P. Van Mieghem, J. Omic, and R. Kooij, “Virus spread in networks,” *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 1–14, 2009.
- [32] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, “Catastrophic cascade of failures in interdependent networks,” *Nature*, vol. 464, pp. 1025–1028, 2010.
- [33] J. F. Donges, H. C. Schultz, N. Marwan, Y. Zou, and J. Kurths, “Investigating the topology of interacting networks,” *The European Physical Journal B*, vol. 84, no. 4, pp. 635–651, 2011.
- [34] H. Wang, Q. Li, G. DiAgostino, S. Havlin, H. E. Stanley, and P. Van Mieghem, “Effect of the interconnected network structure on the epidemic threshold,” *Physical Review E*, vol. 88, no. 2, p. 022801, 2013.
- [35] F. Radicchi and A. Arenas, “Abrupt transition in the structural formation of interconnected networks,” *Nature Physics*, 2013.
- [36] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, “Robustness of interdependent networks under targeted attack,” *Physical Review E*, vol. 83, no. 6, p. 065101, 2011.
- [37] F. Chung, L. Lu, and V. Vu, “Spectra of random graphs with given expected degrees,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 11, pp. 6313–6318, 2003.
- [38] J. G. Restrepo, E. Ott, and B. R. Hunt, “Approximating the largest eigenvalue of network adjacency matrices,” *Physical Review E*, vol. 76, p. 056119, 2007.
- [39] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, “Epidemic outbreaks in complex heterogeneous networks,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 26, no. 4, pp. 521–529, 2002.
- [40] P. Van Mieghem, H. Wang, X. Ge, S. Tang, and F. A. Kuipers, “Influence of assortativity and degree-preserving rewiring on the spectra of networks,” *The European Physical Journal B*, vol. 76, no. 4, pp. 643–652, 2010.
- [41] W. de Haan, Y. A. Pijnenburg, R. L. Strijers, Y. van der Made, W. M. van der Flier, P. Scheltens, and C. J. Stam, “Functional neural network analysis in frontotemporal dementia and alzheimer’s disease using eeg and graph theory,” *BMC neuroscience*, vol. 10, no. 1, p. 101, 2009.

- [42] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, "Characterization of complex networks: A survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.
- [43] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [44] J. G. Restrepo, E. Ott, and B. R. Hunt, "Onset of synchronization in large networks of coupled oscillators," *Physical Review E*, vol. 71, no. 3, p. 036151, 2005.
- [45] D. J. Klein and M. Randić, "Resistance distance," *Journal of Mathematical Chemistry*, vol. 12, no. 1, pp. 81–95, 1993.
- [46] A. Ghosh, S. Boyd, and A. Saberi, "Minimizing effective resistance of a graph," *SIAM review*, vol. 50, no. 1, pp. 37–66, 2008.
- [47] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
- [48] A. Jamakovic and P. Van Mieghem, "On the robustness of complex networks by using the algebraic connectivity," in *NETWORKING 2008 Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet*, pp. 183–194, Springer, 2008.
- [49] L. Donetti, F. Neri, and M. A. Muñoz, "Optimal network topologies: Expanders, cages, ramanujan graphs, entangled networks and all that," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 08, p. P08007, 2006.
- [50] T. Watanabe and N. Masuda, "Enhancing the spectral gap of networks by node removal," *Physical Review E*, vol. 82, no. 4, p. 046102, 2010.
- [51] M. Barahona and L. M. Pecora, "Synchronization in small-world systems," *Physical review letters*, vol. 89, no. 5, p. 054101, 2002.
- [52] T. Nishikawa, A. E. Motter, Y.-C. Lai, and F. C. Hoppensteadt, "Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize?," *Physical review letters*, vol. 91, no. 1, p. 014101, 2003.
- [53] J. M. Anthonisse, "The rush in a directed graph," *Stichting Mathematisch Centrum. Mathematische Besliskunde*, no. BN 9/71, pp. 1–10, 1971.
- [54] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.
- [55] H. Wang, J. M. Hernandez, and P. Van Mieghem, "Betweenness centrality in a weighted network," *Physical Review E*, vol. 77, no. 4, p. 046105, 2008.
- [56] D. Koschützki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski, "Centrality indices," in *Network analysis*, pp. 16–61, Springer, 2005.

- [57] H.-W. Ma and A.-P. Zeng, “The connectivity structure, giant strong component and centrality of metabolic networks,” *Bioinformatics*, vol. 19, no. 11, pp. 1423–1430, 2003.
- [58] S. B. Seidman, “Network structure and minimum degree,” *Social networks*, vol. 5, no. 3, pp. 269–287, 1983.
- [59] B. Pittel, J. Spencer, and N. Wormald, “Sudden emergence of a giant k -core in a random graph,” *Journal of Combinatorial Theory, Series B*, vol. 67, no. 1, pp. 111–151, 1996.
- [60] S. H. Strogatz, “Exploring complex networks,” *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [61] H. Kesten, “The critical probability of bond percolation on the square lattice equals $1/2$,” *Communications in mathematical physics*, vol. 74, no. 1, pp. 41–59, 1980.
- [62] R. T. Smythe and J. C. Wierman, *First-passage percolation on the square lattice*. Springer-Verlag Berlin: Springer, 1978.
- [63] D. J. A. Welsh, “Percolation and related topics,” *Science progress*, vol. 64, pp. 65–83, 1977.
- [64] T. Ohira and R. Sawatari, “Phase transition in a computer network traffic model,” *Physical Review E*, vol. 58, no. 1, p. 193, 1998.
- [65] M. E. J. Newman and D. J. Watts, “Renormalization group analysis of the small-world network model,” *Physics Letters A*, vol. 263, no. 4, pp. 341–346, 1999.
- [66] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and P. H. Scheltens, “Small-world networks and functional connectivity in alzheimer’s disease,” *Cerebral Cortex*, vol. 17, no. 1, pp. 92–99, 2007.
- [67] Y. Kuramoto, *Chemical oscillations, waves, and turbulence*. Springer, Berlin, 1984.
- [68] A. Barrat and M. Weigt, “On the properties of small-world network models,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 13, no. 3, pp. 547–560, 2000.
- [69] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics reports*, vol. 424, no. 4, pp. 175–308, 2006.
- [70] C. Li, H. Wang, and P. Van Mieghem, “Epidemic threshold in directed networks,” *Physical Review E*, vol. 88, no. 6, p. 062802, 2013.
- [71] S. N. Dorogovtsev and J. F. F. Mendes, “Evolution of networks,” *Advances in physics*, vol. 51, no. 4, pp. 1079–1187, 2002.
- [72] D. J. Daley and J. Gani, *Epidemic modelling: an introduction*. Cambridge University Press, Cambridge, U.K., 2001.

- [73] R. Pastor-Satorras and A. Vespignani, “Epidemic dynamics and endemic states in complex networks,” *Physical Review E*, vol. 63, no. 6, p. 066117, 2001.
- [74] C. Castellano, S. Fortunato, and V. Loreto, “Statistical physics of social dynamics,” *Reviews of modern physics*, vol. 81, no. 2, p. 591, 2009.
- [75] J. Shao, S. Havlin, and H. E. Stanley, “Dynamic opinion model and invasion percolation,” *Physical review letters*, vol. 103, no. 1, p. 018701, 2009.
- [76] Q. Li, L. A. Braunstein, S. Havlin, and H. E. Stanley, “Strategy of competition between two groups based on an inflexible contrarian opinion model,” *Physical Review E*, vol. 84, no. 6, p. 066101, 2011.
- [77] W. Li, A. Bashan, S. V. Buldyrev, H. E. Stanley, and S. Havlin, “Cascading failures in interdependent lattice networks: The critical role of the length of dependency links,” *Physical review letters*, vol. 108, no. 22, p. 228702, 2012.
- [78] X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, “Cascading failures in bipartite graphs: model for systemic risk propagation,” *Scientific reports*, vol. 3, no. 1219, 2013.
- [79] N. T. J. Bailey, *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company, 2nd edition, 1975.
- [80] K. Sznajd-Weron and J. Sznajd, “Opinion evolution in closed community,” *International Journal of Modern Physics C*, vol. 11, no. 06, pp. 1157–1165, 2000.
- [81] R. A. Holley and T. M. Liggett, “Ergodic theorems for weakly interacting infinite systems and the voter model,” *The annals of probability*, pp. 643–663, 1975.
- [82] S. Galam, “Minority opinion spreading in random geometry,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 25, no. 4, pp. 403–406, 2002.
- [83] Y. Li, Y. Liu, J. Li, W. Qin, K. Li, C. Yu, and T. Jiang, “Brain anatomical network and intelligence,” *PLoS computational biology*, vol. 5, no. 5, p. e1000395, 2009.
- [84] M. P. van den Heuvel, C. J. Stam, R. S. Kahn, and H. E. Pol, Hulshoff Pol, “Efficiency of functional brain networks and intellectual performance,” *The Journal of Neuroscience*, vol. 29, no. 23, pp. 7619–7624, 2009.
- [85] F. Bartolomei, I. Bosma, M. Klein, J. C. Baayen, J. C. Reijneveld, T. J. Postma, J. J. Heimans, B. W. van Dijk, J. C. de Munck, A. de Jongh, C. K. S., and S. C. J., “Disturbed functional connectivity in brain tumour patients: evaluation by graph analysis of synchronization matrices,” *Clinical neurophysiology*, vol. 117, no. 9, pp. 2039–2049, 2006.
- [86] M. Chavez, M. Valencia, V. Navarro, V. Latora, and J. Martinerie, “Functional modularity of background activities in normal and epileptic brain networks,” *Physical review letters*, vol. 104, no. 11, p. 118701, 2010.
- [87] S. L. Simpson, S. Hayasaka, and P. J. Laurienti, “Exponential random graph modeling for complex brain networks,” *PLoS One*, vol. 6, no. 5, p. e20039, 2011.

- [88] A. Jamakovic and S. Uhlig, “On the relationships between topological measures in real-world networks,” *Networks and Heterogeneous Media*, vol. 3, no. 2, p. 345, 2008.
- [89] B. C. van Wijk, C. J. Stam, and A. Daffertshofer, “Comparing brain networks of different size and connectivity density using graph theory,” *PLoS One*, vol. 5, no. 10, p. e13701, 2010.
- [90] C. Li, H. Wang, W. de Haan, C. J. 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*, vol. 2011, no. 11, p. P11018, 2011.
- [91] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.
- [92] A. Aertsen, G. Gerstein, M. Habib, and G. Palm, “Dynamics of neuronal firing correlation: modulation of effective connectivity,” *Journal of neurophysiology*, vol. 61, no. 5, 1989.
- [93] T. Montez, K. Linkenkaer-Hansen, B. W. Van Dijk, and C. J. Stam, “Synchronization likelihood with explicit time-frequency priors,” *Neuroimage*, vol. 33, no. 4, pp. 1117–1125, 2006.
- [94] C. J. Stam and B. W. van Dijk, “Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets,” *Physica D: Nonlinear Phenomena*, vol. 163, no. 3, pp. 236–251, 2002.
- [95] W. de Haan, W. M. van der Flier, T. Koene, L. Smits, P. Scheltens, and C. J. Stam, “Disrupted modular brain dynamics reflect cognitive dysfunction in alzheimer’s disease,” *Neuroimage*, vol. 59, no. 4, pp. 3085–3093, 2012.
- [96] C. H. Comin and L. da Fontoura Costa, “Identifying the starting point of a spreading process in complex networks,” *Physical Review E*, vol. 84, no. 5, p. 056105, 2011.
- [97] J. Borge-Holthoefer and Y. Moreno, “Absence of influential spreaders in rumor dynamics,” *Physical Review E*, vol. 85, no. 2, p. 026116, 2012.
- [98] R. Pastor-Satorras and A. Vespignani, “Immunization of complex networks,” *Physical Review E*, vol. 65, no. 3, p. 036104, 2002.
- [99] S. P. Borgatti, “Centrality and network flow,” *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [100] N. E. Friedkin, “Theoretical foundations for centrality measures,” *American journal of Sociology*, vol. 96, pp. 1478–1504, 1991.
- [101] B. Mullen, C. Johnson, and E. Salas, “Effects of communication network structure: Components of positional centrality,” *Social Networks*, vol. 13, no. 2, pp. 169–185, 1991.
- [102] M. E. J. Newman, “The mathematics of networks,” *The new palgrave encyclopedia of economics*, vol. 2, pp. 1–12, 2008.

- [103] P. Van Mieghem, “Graph eigenvectors, fundamental weights and centrality metrics for nodes in networks,” *arXiv preprint arXiv:1401.4580*, 2014.
- [104] P.-J. Kim and H. Jeong, “Reliability of rank order in sampled networks,” *The European Physical Journal B*, vol. 55, no. 1, pp. 109–114, 2007.
- [105] D. Koschützki and F. Schreiber, “Comparison of centralities for biological networks,” in *German Conference on Bioinformatics*, pp. 199–206, 2004.
- [106] E. Estrada, “Characterization of topological keystone species: local, global and meso-scale centralities in food webs,” *Ecological Complexity*, vol. 4, no. 1, pp. 48–57, 2007.
- [107] C. Li, H. Wang, and P. Van Mieghem, “Degree and principal eigenvectors in complex networks,” in *Proceedings of NETWORKING 2012*, pp. 149–160, Springer, 2012.
- [108] P. Van Mieghem, D. Stevanović, F. A. 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*, vol. 84, no. 1, p. 016101, 2011.
- [109] S. Galam, “Local dynamics vs. social mechanisms: A unifying frame,” *EPL (Europhysics Letters)*, vol. 70, no. 6, p. 705, 2005.
- [110] Q. Li, L. A. Braunstein, H. Wang, J. Shao, H. E. Stanley, and S. Havlin, “Non-consensus opinion models on complex networks,” *Journal of Statistical Physics*, vol. 151, no. 1-2, pp. 92–112, 2013.
- [111] B. Qu, Q. Li, S. Havlin, H. E. Stanley, and H. Wang, “Non-consensus opinion model on directed networks,” *arXiv preprint arXiv:1404.7318*, 2014.
- [112] M. P. Joy, A. Brock, D. E. Ingber, and S. Huang, “High-betweenness proteins in the yeast protein interaction network,” *BioMed Research International*, vol. 2005, no. 2, pp. 96–103, 2005.
- [113] S. Trajanovski, J. Martín-Hernández, W. Winterbach, and P. Van Mieghem, “Robustness envelopes of networks,” *Journal of Complex Networks*, vol. 1, no. 1, pp. 44–62, 2013.
- [114] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Random graphs with arbitrary degree distributions and their applications,” *Physical Review E*, vol. 64, no. 2, p. 026118, 2001.
- [115] M. Krivelevich and B. Sudakov, “The largest eigenvalue of sparse random graphs,” *Combinatorics, Probability and Computing*, vol. 12, no. 1, pp. 61–72, 2003.
- [116] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek, “Spectra of real-world graphs: Beyond the semicircle law,” *Physical Review E*, vol. 64, no. 2, p. 026704, 2001.
- [117] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control*, vol. 28. Oxford University Press, Oxford, U.K., 1991.
- [118] C. Castellano and R. Pastor-Satorras, “Thresholds for epidemic spreading in networks,” *Physical review letters*, vol. 105, no. 21, p. 218701, 2010.

- [119] J. O. Kephart and S. R. White, “Directed-graph epidemiological models of computer viruses,” in *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 343–359, IEEE, 1991.
- [120] P. Van Mieghem, “Epidemic phase transition of the sis type in networks,” *EPL (Europhysics Letters)*, vol. 97, no. 4, p. 48004, 2012.
- [121] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, “Discrete-time markov chain approach to contact-based disease spreading in complex networks,” *EPL (Europhysics Letters)*, vol. 89, no. 3, p. 38009, 2010.
- [122] E. Cator and P. Van Mieghem, “Second-order mean-field susceptible-infected-susceptible epidemic threshold,” *Physical Review E*, vol. 85, no. 5, p. 056111, 2012.
- [123] 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 B: Biological Sciences*, vol. 277, no. 1701, pp. 3827–3835, 2010.
- [124] H. Hinrichsen, “Non-equilibrium critical phenomena and phase transitions into absorbing states,” *Advances in physics*, vol. 49, no. 7, pp. 815–958, 2000.
- [125] P. Van Mieghem and E. Cator, “Epidemics in networks with nodal self-infection and the epidemic threshold,” *Physical Review E*, vol. 86, no. 1, p. 016116, 2012.
- [126] M. Boguñá and R. Pastor-Satorras, “Epidemic spreading in correlated complex networks,” *Physical Review E*, vol. 66, no. 4, p. 047104, 2002.
- [127] P. Van Mieghem, “The n-intertwined sis epidemic network model,” *Computing*, vol. 93, no. 2-4, pp. 147–169, 2011.
- [128] P. Van Mieghem, “The viral conductance of a network,” *Computer Communications*, vol. 35, no. 12, pp. 1494–1506, 2012.
- [129] C. Li, H. Wang, and P. Van Mieghem, “Bounds for the spectral radius of a graph when nodes are removed,” *Linear Algebra and its Applications*, vol. 437, no. 1, pp. 319–323, 2012.
- [130] S. M. Cioabă, “A necessary and sufficient eigenvector condition for a connected graph to be bipartite,” *Electron. J. Linear Algebra (ELA)*, vol. 20, pp. 351–353, 2010.
- [131] J. G. Restrepo, E. Ott, and B. R. Hunt, “Characterizing the dynamical importance of network nodes and links,” *Physical Review Letters*, vol. 97, no. 9, p. 094102, 2006.
- [132] A. Milanese, J. Sun, and T. Nishikawa, “Approximating spectral impact of structural perturbations in large networks,” *Physical Review E*, vol. 81, no. 4, p. 046112, 2010.
- [133] M. R. Garey and D. S. Johnson, “Computer and intractability,” *A Guide to the NP-Completeness*. Ney York, NY: WH Freeman and Company, 1979.

- [134] J. Omic, J. Martín-Hernández, and P. Van Mieghem, “Network protection against worms and cascading failures using modularity partitioning,” in *Proceedings of 22nd International Teletraffic Congress (ITC 22), September 7-9, Amsterdam, Netherlands*, pp. 1–8, IEEE, 2010.
- [135] D. Taylor and J. G. Restrepo, “Modifying network connectivity with a subgraph addition,” *arXiv preprint: 1102.4876v1*, 2011.
- [136] T. Nishikawa and A. E. Motter, “Network synchronization landscape reveals compensatory structures, quantization, and the positive effect of negative interactions,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 23, pp. 10342–10347, 2010.
- [137] D. Cvetković, M. Doob, and H. Sachs, “Spectra of graphs: Theory and applications,” *Johann Ambrosius Barth, 3rd Edition, Heidelberg, Leipzig*, 1995.
- [138] N. Robertson and P. D. Seymour, “Graph minors. xiii. the disjoint paths problem,” *Journal of Combinatorial Theory, Series B*, vol. 63, no. 1, pp. 65–110, 1995.
- [139] H. Yuan, “A bound on the spectral radius of graphs,” *Linear Algebra and its Applications*, vol. 108, pp. 135–139, 1988.
- [140] N. D. Roussopoulos, “A max {m, n} algorithm for determining the graph h from its line graph g,” *Information Processing Letters*, vol. 2, no. 4, pp. 108–112, 1973.
- [141] M. Aouchiche, F. K. Bell, D. Cvetković, P. Hansen, P. Rowlinson, S. Simić, and D. Stevanović, “Variable neighborhood search for extremal graphs. 16. some conjectures related to the largest eigenvalue of a graph,” *European Journal of Operational Research*, vol. 191, no. 3, pp. 661–676, 2008.
- [142] D. Stevanović, “The largest eigenvalue of nonregular graphs,” *Journal of combinatorial theory, Series B*, vol. 91, no. 1, pp. 143–146, 2004.
- [143] K.-I. Goh, B. Kahng, and D. Kim, “Spectra and eigenvectors of scale-free networks,” *Physical Review E*, vol. 64, no. 5, p. 051903, 2001.
- [144] E. N. Gilbert, “Random graphs,” *The Annals of Mathematical Statistics*, pp. 1141–1144, 1959.
- [145] B. Bollobás, “A probabilistic proof of an asymptotic formula for the number of labelled regular graphs,” *European Journal of Combinatorics*, vol. 1, no. 4, pp. 311–316, 1980.
- [146] M. Molloy and B. Reed, “A critical point for random graphs with a given degree sequence,” *Random structures and algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.
- [147] P. Boldi and S. Vigna, “The webgraph framework i: compression techniques,” in *Proceedings of the 13th international conference on World Wide Web*, pp. 595–602, ACM, 2004.

- [148] P. Boldi, M. Rosa, M. Santini, and S. Vigna, “Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks,” in *Proceedings of the 20th international conference on World Wide Web*, pp. 587–596, ACM, 2011.
- [149] G. Bianconi, N. Gulbahce, and A. E. Motter, “Local structure of directed networks,” *Physical review letters*, vol. 100, no. 11, p. 118701, 2008.
- [150] G. Robins, P. Pattison, and P. Wang, “Closure, connectivity and degree distributions: Exponential random graph (p^*) models for directed social networks,” *Social Networks*, vol. 31, no. 2, pp. 105–117, 2009.
- [151] Y. Kim, S.-W. Son, and H. Jeong, “Finding communities in directed networks,” *Physical Review E*, vol. 81, no. 1, p. 016103, 2010.
- [152] R. van den Brink and R. P. Gilles, “Measuring domination in directed networks,” *Social Networks*, vol. 22, no. 2, pp. 141–157, 2000.
- [153] M. E. J. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
- [154] D. Garlaschelli and M. I. Loffredo, “Patterns of link reciprocity in directed networks,” *Physical Review Letters*, vol. 93, no. 26, p. 268701, 2004.
- [155] S. M. Park and B. J. Kim, “Dynamic behaviors in directed networks,” *Physical Review E*, vol. 74, no. 2, p. 026114, 2006.
- [156] N. Schwartz, R. Cohen, D. Ben-Avraham, A.-L. Barabási, and S. Havlin, “Percolation in directed scale-free networks,” *Physical Review E*, vol. 66, no. 1, p. 015104, 2002.
- [157] M. E. J. Newman, S. Forrest, and J. Balthrop, “Email networks and the spread of computer viruses,” *Physical Review E*, vol. 66, no. 3, p. 035101, 2002.
- [158] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, “Network robustness and fragility: Percolation on random graphs,” *Physical review letters*, vol. 85, no. 25, p. 5468, 2000.
- [159] M. Boguñá and M. Á. Serrano, “Generalized percolation in random directed networks,” *Physical Review E*, vol. 72, no. 1, p. 016106, 2005.
- [160] L. A. Meyers, M. E. J. Newman, and B. Pourbohloul, “Predicting epidemics on directed contact networks,” *Journal of theoretical biology*, vol. 240, no. 3, pp. 400–418, 2006.
- [161] 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*, vol. 110, no. 10, p. 108701, 2013.
- [162] N. Chen and M. Olvera-Cravioto, “Directed random graphs with given degree distributions,” *Stochastic Systems*, vol. 3, no. 1, p. 147, 2013.

- [163] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.
- [164] F. Juhász, "On the characteristic values of non-symmetric block random matrices," *Journal of Theoretical Probability*, vol. 3, no. 2, pp. 199–205, 1990.
- [165] P. Bonacich, "Power and centrality: A family of measures," *American journal of sociology*, pp. 1170–1182, 1987.
- [166] M. T. Thai and P. M. Pardalos, *Handbook of optimization in complex networks*. Springer, 2012.
- [167] R. B. Bapat, J. W. Grossman, and D. M. Kulkarni, "Generalized matrix tree theorem for mixed graphs," *Linear and Multilinear Algebra*, vol. 46, no. 4, pp. 299–312, 1999.
- [168] P. Van Mieghem, G. Hooghiemstra, and R. W. van der Hofstad, "Stochastic model for the number of traversed routers in internet," in *Proceedings of Passive and Active Measurement (PAM), RIPE NCC*, 2001.

Samenvatting (Dutch Summary)

HEB je enig idee hoeveel netwerken je gebruikt wanneer je je Facebook of Twitter checkt in de trein? Je hersen-netwerk bepaalt je gedrag en je metabolisch netwerk zorgt voor je energie. Je vormt een sociaal netwerk met je medereizigers. Pas op, misschien verspreidt zich wel een virus over het sociale netwerk door de coupe! Wanneer je je telefoon of laptop gebruikt om Facebook of Twitter te bezoeken zijn ze verbonden met het Internet, je gebruikersprofielen staan op het WereldWijde Web (WWW), en elektriciteit loopt via het energienetwerk naar je apparaten. Bovendien worden Facebook en Twitter gezien als een online sociaalnetwerk. Nu gaat je telefoon, inderdaad, je bent verbonden met een communicatienetwerk! Maar wacht je zit in de trein, en die maakt deel uit van het transportnetwerk. Het is ongelofelijk, maar complexe netwerken zijn onmisbaar geworden in ons dagelijks leven. Dit proefschrift begint met een introductie in de geschiedenis van complexe netwerken als onderzoeksveld en de benodigde achtergrondkennis, zoals netwerk kengetallen, netwerk modellen en dynamische processen. Gemotiveerd door een beter begrip van echte netwerken, verkennen onderzoekers de eigenschappen die gezien worden in echte netwerken. Een veelvoud aan kengetallen zijn voorgesteld om de eigenschappen van netwerken te kwantificeren, en verder te karakteriseren.

Een essentiële vraag die opkomt is: “*Hoe kunnen we netwerken beter en efficiënter karakteriseren?*” We kunnen een netwerk bijvoorbeeld karakteriseren vanuit verschillende aspecten: of het netwerk goed geconnecteerd is, of de knopen in het netwerk dicht bij elkaar liggen, of de burenen van een knoop ook onderling burenen zijn, of knopen verbonden zijn met andere knopen met soortgelijke graad, of het netwerk vatbaar is voor een virusuitbraak, enzovoort. Als we meerdere kengetallen gebruiken om een netwerk te karakteriseren, is er dan sprake van overtollige informatie? Als we, daarentegen, maar één kengetal gebruiken, is dat dan voldoende om het netwerk volledig te beschrijven? Deel I van dit proefschrift gaat in op het selecteren van een set representatieve kengetallen die samen efficiënt en adequaat een netwerk karakteriseren. Het patroon van correlaties tussen de kengetallen zoals bestudeerd in hersennetwerken blijkt consistent te zijn met dat wat er voor Erdős-Rényi grafen gevonden is. Naast het karakteriseren van een netwerk als geheel, is het belangrijk om de knopen in een netwerk te kwantificeren. Als bijvoor-

beeld twee bedrijven elkaar beconcurreren om klanten, zullen ze proberen invloedrijke klanten voor hun producten te winnen. Als de twee bedrijven verschillende marketing strategieën gebruiken om invloedrijke klanten te selecteren, zal de uitkomst van de strijd anders zijn. Welke klanten zouden geselecteerd moeten worden als “Very Important Persons (VIP)”: klanten die veel vrienden hebben, klanten met een grote reputatie, klanten met beïnvloedbare vrienden, of simpelweg willekeurige klanten? In deel I worden ook kengetallen bestudeerd die knopen rangschikken naar mate van hoe centraal in het netwerk ze liggen. De correlaties tussen deze kengetallen wordt beschreven in netwerk modellen en echte netwerken. Daarbij passen we centraliteit-kengetallen toe in een opinie dynamisch proces om één van de twee opinies de competitie tussen twee groepen te helpen winnen. We tonen aan dat de knopen in een netwerk efficiënt gekarakteriseerd kunnen worden met behulp van een eenvoudig kengetal als benadering voor een complex centraliteit-kengetal.

Een beter inzicht in de eigenschappen van een netwerk kan leiden tot een beter begrip van het dynamische en functionele gedrag. We vergelijken twee Susceptible-Infected-Susceptible (SIS) gemiddeldveldbenaderingen aan de hand van een ε -SIS model als maatstaf. We tonen aan dat de N-intertwined mean-field approximation (NIMFA) benadering de betere is. Een spectraal kengetal, het spectrale bereik, wordt verondersteld een betere kwantificatie te zijn van hoe weerbaar een netwerk is tegen virusverspreiding dan de graaddiversiteit. Deel II richt zich met name op hoe we netwerken minder gevoelig voor virussen kunnen maken. Door het spectrale bereik te minimaliseren zou de epidemische drempel verhoogd kunnen worden. In deel II worden verschillende manieren om de epidemische drempel te verhogen gepresenteerd. Het minimaliseren van het spectrale bereik door middel van het verwijderen van knopen of zijden is een NP-hard probleem. In de meeste gevallen is de aanpak waarbij de zijden met het grootste product van de voornaamste eigenvector componenten verwijderd worden superieur aan andere methodes. We stellen limieten aan de afname van het spectrale bereik als gevolg van het verwijderen van knopen of zijden. Tot op heden zijn epidemieën in netwerken vooral bestudeerd in ongerichte netwerken. Echter, veel echte netwerken, zoals online sociale netwerken en het WWW, waarover informatie, emoties, of malware zich verspreidt zijn gerichte netwerken. In deel II worden twee algoritmen voorgesteld om netwerken te genereren met een bepaalde gerichtheid. Wij tonen aan dat de epidemische drempel verhoogd kan worden door de gerichtheid van een netwerk te vergroten via het verleggen van zijden of het resetten ervan.

Tenslotte trekken we de hoofdconclusies, geven een opsomming van de bijdragen in dit proefschrift, en stellen enkele onderzoeksproblemen.

Cong Li
Delft, april 2014

Acknowledgements

I would like to express my deepest gratitude to my promoter Prof. Piet Van Mieghem and copromoter Dr. Huijuan Wang, for their endless patience, kind help and countless times of modifying my scientific papers. It is my honor to be their PhD student. Their enthusiasm and academic rigor for the research encourage me a lot. With their guidance, I clearly see the improvement of my paper writing and my presentation skills. I am also grateful to get their friendly help for my life problem other than work. Thank you very much and I will miss you two.

This thesis would not have been possible without the help as well as presence of many people, including but not limited to: my husband and our parents for their infinite love; my Chinese professor Fu Liu for encouraging me to travel abroad; all committee members for their precious time and valuable feedback; all my roommates for giving me a comfortable home and last but not least, my dearest office mates, Ruud van de Bovenkamp, Javier Martín Hernández, Marcus Märtens, Anteneh Ayalew Beshir, Dongchao Guo for the kindness in their hearts.

Before this page runs too long, I want to mention: a special group of people from NAS—Edgar, Fernando, Rob, Norbert, Stojan, Jil, Wendy, Niels, Song, Farabi, Rogier, Christian, Nico, Remco, Hale, Negar, Yakup, Shruti, Ebisa, Dajie, Wynand, Jasmina, Siyu, Yue, António, Annalisa, Evangelos; my good friends in Delft—Tianduowa Zhu, Bo Qu, Xiangrong Wang, Xiaoxi Wang, Yao Gao, Jing Zhao, Chang Yu, Kun Yan, Zheng Xu, Shuang Zhang, Bing Liu, Yue Chen and Shuhong Tan; all my friends in China who give me mental supports.

To all of You, **thanks!**

Cong Li
Delft, May 2014

Publications by the Author

- P.1** C. Li, Q. Li, P. Van Mieghem, H. Eugene Stanley and H. Wang, “Correlation Between Centrality Metrics and Their Application to the Opinion Model”, *submitted to European Physical Journal B*, 2014.
- P.2** C. Li, H. Wang, P. Van Mieghem, “New lower bounds for the fundamental weight of the principal eigenvector in complex networks”, *submitted to Third International Workshop on Complex Networks and their Applications*, 2014. [Not Related to this thesis]
- P.3** C. Li, H. Wang and P. Van Mieghem, “Epidemic Threshold in Directed Networks”, *Physical Review E*, 88(6), 062802, 2013.
- P.4** C. Li, R. van de Bovenkamp and P. Van Mieghem, “Susceptible-Infected-Susceptible Model: A Comparison of N-intertwined and Heterogeneous Mean-field Approximations”, *Physical Review E*, 86(2), 026116, 2012.
- P.5** C. Li, H. Wang and P. Van Mieghem, “Bounds for the Spectral Radius of a Graph When Nodes are Removed”, *Linear Algebra and its Applications*, 437, pp. 319-323, 2012.
- P.6** C. Li, H. Wang and P. Van Mieghem, “Degree and Principal Eigenvectors in Complex Networks”, *Proceedings of IFIP Networking Conference*, Part I, LNCS 7289, pp. 149-160, Prague, Czech Republic, 2012.
- P.7** C. Li, H. Wang, W. de Haan, C. J. 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 (JSTAT)*, 11018, 2011.
- P.8** P. Van Mieghem, D. Stevanovic, F. A. 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(1), 016101, 2011.

P.9 C. Li, F. Liu and Y. Zhang, “A Principal Palm-line Extraction Method for Palmprint Images Based on the Diversity and Contrast”, *Processing of the 3rd International Congress on Image and Signal (CISP)*, IEEE, vol. 4, pp. 1772-1777, Yantai, China, 2010. [Not related to this thesis]

P.10 L. Sun, C. Li, D. Yang and Q. Wang, “Development of Automatic Control System of Spandex Production and Total”, *Journal of JiLin University (information Science Edition)*, 28(05), pp. 011, 2010. [Not Related to this thesis]

TU Delft Technical Reports

R.1 J. Martin Hernandez, S. Trajanovski, H. Wang, C. Li and P. Van Mieghem, “Zero and non-zero eigenvector components of graph matrices”, *Delft University of Technology*, report20121109, 2012. [Not Related to this thesis]

Relations to this Thesis

This thesis consists of previously published publications by the present author. The following table E.1 provides a relation between the afore-mentioned publications and chapters of this thesis.

Publication	Intro.	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch.6	Ch. 7	Ch. 8
P.1		○		●				
P.3		○					●	
P.4		○			●			
P.5						●		
P.6		○		●		○		
P.7		○	●					
P.8						●		

Table E.1: Relation between chapters of the thesis and the list of publications by the author. The meaning of the bullets is: ● major relation, ○ minor relation

Curriculum Vitae



Cong Li was born in Harbin (China) on 2nd August, 1986. In September 2009, she graduated in China from Jilin University (JLU), with a B.Sc. degree in Telecommunication Engineering. She was recommended as a postgraduate candidate in Pattern Recognition and Intelligent System exempt from admission exam. Thanks to the China program CSC, she was awarded a scholarship to pursue her PhD study abroad, in The Netherlands. She became a PhD candidate in the Networks Architectures and Services group, faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology (TU Delft), in September 2010.

During her PhD period, Cong focused on the analysis and structural design of complex networks and published 6 papers in peer-reviewed international journals and conferences, under the supervision of Prof. P. Van Mieghem and the daily supervision dr. Huijuan Wang. Her present research interests are in the field of network properties, dynamic processes and functional brain networks. She assisted Prof. P. Van Mieghem at teaching, supervising and grading the lectures, exam and homework for an M.Sc course, the Performance Analysis, and has reviewed several international journal and conference papers.

物格而後知至
知至而後意識
意識而後心正
心正而後身修
身修而後家齊
家齊而後國治
國治而後天下平

語出大學書于甲午馬年

Things being investigated, knowledge became complete.
Their knowledge being complete, their thoughts were sincere.
Their thoughts being sincere, their hearts were then rectified.
Their hearts being rectified, their persons were cultivated.
Their persons being cultivated, their families were regulated.
Their families being regulated, their states were rightly governed.
Their states being rightly governed, the entire world was at peace.

— The Great Learning
(c.206 B.C.- c.9 A.D. in Han Dynasty)

