Robustness of complex networks

Theory and application

Wang, Xiangrong

DOI
10.4233/uuid:c107cc92-d275-45df-ad56-b754e8ead98c

Publication date
2016

Document Version
Publisher's PDF, also known as Version of record

ROBUSTNESS OF COMPLEX NETWORKS: THEORY AND APPLICATION

ROBUSTNESS OF COMPLEX NETWORKS: THEORY AND APPLICATION

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 woensdag 21 december 2016 om 10:00 uur

door

Xiangrong WANG

Master of Science in Electrical Engineering van
Beijing Jiaotong University, Beijing, China
geboren te Mengyin, Shandong Provincie, China.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. P. F. A. Van Mieghem
promotor: prof. dr. ir. R. E. Kooij

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. P. F. A. Van Mieghem,	Technische Universiteit Delft, promotor
Prof. dr. ir. R. E. Kooij,	Technische Universiteit Delft and TNO, promotor

Onafhankelijke leden:

Prof. dr. Y. Moreno	University of Zaragoza
Prof. dr. J. L Marzo	Universitat de Girona
Prof. dr. P. M. A. Sloot	University of Amsterdam
Prof. dr. ir. D. H. J. Epema	Technische Universiteit Delft
Dr. O. Cats	Technische Universiteit Delft
Prof. dr. C. Witteveen	Technische Universiteit Delft, reservelid

This research was supported by the China Scholarship Council.

Keywords: Complex Networks, Robustness of Networks, Graph Spectra, Power Grids, Metro Networks, Line Graph, Eigenvectors/Eigenvalues, Interdependent Networks

Author email: xiangrongwang88@gmail.com

ISBN 978-94-6186-775-9

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

To my parents

SUMMARY

Failures of networks, such as power outages in power systems, congestions in transportation networks, paralyse our daily life and introduce a tremendous cascading effect on our society. Networks should be constructed and operated in a robust way against random failures or deliberate attacks.

We study how to add a single link into an existing network such that the robustness of the network is maximally improved among all the possibilities. A graph metric, the effective graph resistance, is employed to quantify the robustness of the network. Though exhaustive search guarantees the optimal solution, the computational complexity is high and is not scalable with the increase of network size. We propose strategies that take into account the structural and spectral properties of networks and indicate links whose addition result in a high robustness level.

To apply the effective graph resistance to real-world power grids and to cope with the robustness of dynamical processes, we improve the robustness of power grids against cascading failures by adding transmission lines. Compared to the existing robustness metrics investigated in power grids, the effective graph resistance effectively quantifies the robustness by taking into account multiple paths and their ability to accommodate power flows. Experimental results suggest the existence of Braess's paradox in power grids: introducing an additional line into the system occasionally results in the decrease of the grid robustness.

Network science and graph theory are applied to investigate the robustness of 33 worldwide metro networks under random failures or targeted attacks. Ten theoretical and three numerical robustness metrics are studied in the metro networks. We find that the robustness metrics capture two distinct aspects of the robustness of metro networks: (i) several metrics place an emphasis on alternative paths and (ii) other metrics highlight the length of the paths.

Robustness of networks is threatened by link failures in real-world networks, for example failures of transmission lines in power grids. To analyse the robustness

of a network against link failures, we study line graphs which transform links in the original graph into nodes. Fundamental properties including degree distribution, degree assortativity of a line graph are explored. The line graphs of Erdös-Rényi random graphs show the same degree distribution pattern. In addition, we find that most synthetic and real-world networks exhibit positive assortativity in the corresponding line graphs. Meanwhile, we find trees and non-trees consisting of cycles and paths whose line graphs have negative assortativity.

Though various robustness metrics have been proposed and widely studied, the spectrum of graph matrices is hardly understood. We approach the challenge by studying the eigenvector matrix of the Laplacian matrix of a graph. We try to understand fundamental properties of the eigenvector matrix such as number of zeros, the sum of the elements, the maximum and the minimum element. For the particular class of Erdös-Rényi random graphs, we find that a product of a Gaussian and a super-Gaussian distribution approximates accurately the distribution of a randomly chosen component from the row sum of the eigenvector matrix of the Laplacian.

The study of single networks is limited to anticipating the interaction property between real-world networks, particularly between the critical infrastructures. Interdependent networks are proposed by researchers to incorporate the interconnections between different networks.

Modelling the interconnection pattern between networks is a challenge in the study of interdependent networks. Motivated by spatial networks where links between nodes are determined by locations of nodes, we investigate two interconnection topologies, the random geometric graph and the relative neighbourhood graph. The two interconnection topologies generalize the one-to-one interconnection to an arbitrary number of interconnections depending on the locations of nodes. To evaluate the robustness of the two interconnection topologies against node failures, we investigate the impact of node failures on the interdependent network, where the robustness is quantified by the largest mutually connected component. We find that the random geometric graph shows a higher robust level compared to the relative neighbourhood graph. In addition, we propose the derivative of the largest mutually connected component as a new robust metric which addresses the impact of a small fraction of node failures. To avoid the collapse of the whole network, the

proposed robustness metric quantifies the damage of networks triggered by a small fraction of failures, significantly smaller than the fraction at the critical threshold that corresponds to the collapse of the whole network.

Real-world networks, such as smart grids consisting of sensor networks, power networks and coupled infrastructures of power systems and fibre-optic communication systems, show a multiple-to-multiple interconnection pattern, which means that one node in one network connects to multiple nodes in the other network and vice versa. Different from the one-to-one interconnection pattern studied in literature, we study a general regular interconnection pattern (constant row and column sum). Consider an interdependent network consisting of two different types of graphs G_1 and G_2 with the weight p on each interconnection link. If the interconnection matrix $B = pI$, where I is the identity matrix, there exists a structural transition threshold p^*, where dynamic processes are separated into two regimes: (a) $p > p^*$, the network acts as a whole; (b) $p < p^*$, the network operates as if the graph is separated G_1 and G_2. For the interdependent network with a regular interconnection matrix $B \neq pI$, our findings include (i) an upper bound for the transition threshold p^*; (ii) topologies of interdependent networks where the upper bound is reached; (iii) the interpretation of the transition threshold p^* in terms of the minimum cut; (iv) the exact transition threshold p^* for special scenarios; (v) a counter-example to show that the structural transition p^* does not always exist.

SAMENVATTING

Uitval van netwerken, zoals stroomuitval in elektriciteitssystemen, opstoppingen in transportnetwerken, verlammen ons dagelijks leven en introduceren een enorm domino-effect in onze maatschappij. Netwerken moeten op een robuuste manier worden gebouwd en bestuurd tegen een toevallige uitval van functie of opzettelijke aanvallen. We onderzoeken hoe een enkele verbinding in een bestaand netwerk kan worden aangebracht, zodanig dat die van alle mogelijkheden de robuustheid van het netwerk maximaal verbetert. Een parameter voor grafen, de effective graph resistance, is toegepast om de robuustheid van het netwerk te kwantificeren. Hoewel uitputtend onderzoekswerk de optimale oplossing garandeert, is de rekenkundige complexiteit groot en kan niet worden opgeschaald met het toenemen van het netwerkformaat. We stellen strategieën voor, die rekening houden met de structurele en spectrale eigenschappen van netwerken, en wijzen verbindingen aan, waarvan de toevoeging in een hoog niveau van robuustheid resulteert.

Voor de toepassing van de effective graph resistance in real-world elektriciteitsnetwerken en het hanteren van de robuustheid van dynamische processen, verbeteren we de robuustheid van de elektriciteitsnetwerken met de toevoeging van transmissielijnen. Vergeleken met bestaande parameters voor robuustheid, die zijn onderzocht in elektriciteitsnetten, kwantificeert de effective graph resistance de robuustheid door rekening te houden met een verscheidenheid aan paden en hun vermogen elektrische stroom te vervoeren. Experimentele resultaten geven aanwijzingen voor het bestaan van de Braess-paradox in elektriciteitsnetwerken: het aanbrengen in het systeem van een toegevoegde lijn resulteert soms in een afname van de robuustheid van het net.

Netwerkwetenschap en de graaftheorie worden toegepast om de robuustheid van wereldwijd 33 metronetwerken te onderzoeken tijdens een toevallige uitval of gerichte aanvallen. Tien theoretische en 3 numerieke maten voor robuustheid worden onderzocht in de metronetwerken. We vinden dat de maat voor robuustheid

twee verschillende aspecten weergeven van de robuustheid van metro netwerken: (i) diverse maten leggen de nadruk op alternatieve paden en (ii) andere maten benadrukken de lengte van de paden.

De robuustheid van netwerken wordt bedreigd door uitval van verbindingen in real-world netwerken, bijvoorbeeld uitval van transmissielijnen in elektriciteitsnetten. Om de robuustheid van een netwerk tegen uitval van verbindingen te analyseren, onderzoeken we lijngrafen die verbindingen in de oorspronkelijke graaf veranderen in knooppunten. Fundamentele eigenschappen, waaronder de degree verdeling en degree assortativiteit van een lijngraaf, worden onderzocht. De lijngrafen van Erdös-Rényi random grafen laten hetzelfde patroon van degree verdeling zien. Daarnaast vinden we dat de meeste gemodelleerde en real-world netwerken positieve assortativiteit vertonen in de overeenkomstige lijngrafen. Ondertussen vinden we trees en non-trees bestaande uit cycli en paden, waarvan de lijngrafen negatieve assortativiteit hebben.

Hoewel diverse parameters voor robuustheid zijn voorgesteld en uitgebreid bestudeerd, wordt het spectrum van graafmatrixen amper begrepen. We benaderen de uitdaging door de eigenvector matrix van de Laplace matrix van een graaf te onderzoeken. We proberen de fundamentele eigenschappen van de eigenvector matrix te begrijpen, zoals het aantal nullen, de som van de elementen, het maximum en het minimum element. Voor de afzonderlijke klasse van Erdös-Rényi random grafen vinden we dat een product van een Gauss- en een super-Gaussverdeling nauwkeurig de verdeling van een willekeurig gekozen component van de rijensom van de eigenvector matrix van de Laplace matrix benaderen.

Het bestuderen van enkelvoudige netwerken wordt beperkt tot het anticiperen op de eigenschap van interactie tussen real-world netwerken, met name tussen de kritische infrastructuren. Onderling afhankelijke netwerken worden door onderzoekers voorgesteld om de onderlinge verbindingen tussen verschillende netwerken te belichamen.

Een model maken van het interconnectiepatroon van onderlinge verbindingen tussen netwerken is een uitdaging bij het bestuderen van onderling afhankelijke netwerken. Geïnspireerd door ruimtelijke netwerken waarin verbindingen tussen knooppunten bepaald worden door de locatie van de knooppunten, onderzoeken we twee vermaasde topologieën, de random geometric graph en de relative neigh-

bourhood graph. De twee interconnectie topologieën generaliseren de één-op-één interconnectie naar een willekeurig aantal interconnecties afhankelijk van de locatie van de knooppunten. Om de robuustheid van de twee vermaasde topologieën tegen uitval van knooppunten te evalueren, onderzoeken we de uitwerking van uitval van knooppunten op het onderling afhankelijke netwerk, waarvan de robuustheid wordt gekwantificeerd door middel van het grootste gemeenschappelijk verbonden component. We vinden dat de random geometric graph een hoger niveau van robuustheid vertoont in vergelijking met de relative neighbourhood graph. Daarnaast stellen we het afleiden van het grootste gemeenschappelijk verbonden component voor als een nieuwe parameter voor robuustheid, die de uitwerking van een uitval van een klein deel van de knooppunten weergeeft. Om de ineenstorting van het hele netwerk te vermijden, kwantificeert de voorgestelde maat van robuustheid de schade aan netwerken uitgelokt door uitval van een klein deel, significant kleiner dan de hoeveelheid van de kritische drempel waarbij een ineenstorting van het hele netwerk optreedt.

Real-world scenario's, zoals smart grids bestaande uit sensornetwerken, elektriciteitsnetwerken en eraan gekoppeld infrastructuren van elektriciteitssystemen en glasvezelcommunicatiesystemen, vertonen een multiple-to-multiple interconnectiepatroon, wat betekent dat één knooppunt in één netwerk verbonden is met meerdere knooppunten in het andere netwerk en vice versa. Verschillend van het één-op-één interconnectiepatroon, dat in de literatuur bestudeerd wordt, bestuderen wij een geheel regelmatig interconnectiepatroon (gelijke totalen in rijen en kolommen). Overweeg een onderling afhankelijk netwerk bestaande uit twee verschillende vormen van grafen G_1 en G_2 met een gewicht p op elke onderlinge verbinding. Als de interconnectie matrix $B = pI$, waarbij I de identiteitsmatrix is, dan bestaat daar een structurele overgangsdrempel p^*, waarbij dynamische processen worden gescheiden in twee regimes: (a) $p > p^*$, het netwerk werkt als een geheel; (b) $p < p^*$, het netwerk werkt alsof de grafen G_1 and G_2 zijn gescheiden. Voor het onderling afhankelijke netwerk met een regelmatige interconnectie matrix $B \neq pI$, omvatten onze bevindingen (i) een bovengrens aan de overgangsdrempel p^* (ii)topologieën van onderling afhankelijke netwerken waarvan de bovengrens is bereikt (iii) een interpretatie van de overgangsdrempel p^* in termen van de minimum cut (iv) een exacte overgangsdrempel p^* voor speciale scenario's (v)een tegenvoorbeeld om te

tonen dat de structurele overgang p^* niet altijd bestaat.

CONTENTS

1

INTRODUCTION

Networks exist everywhere in the world and in our daily lives. Examples include transportation networks (airline, metro, train and bus networks) [1], power/gas/water networks [2], telecommunication networks [3], the Internet [4], social networks (Facebook, Twitter, LinkedIn) [5], biological networks [6] and so on.

Though a commonly agreed definition for the robustness of networks does not seem to exist, we interpret the robustness of networks, in this dissertation, as the maintenance of functionality under external perturbations such as random failures or targeted attacks. Motivation to study the robustness of networks is that failures of networks affect directly the services running on the networks and introduce tremendous cascading impact on our societies and our daily lives. Worldwide power outages since 1960s, such as United States, India, Brazil [7], are examples of failures of power networks. In August 14, 2003, the power outage in U.S.-Canada affected an area with an estimated 50 million people and an estimate of total costs ranging from 4 billion to 10 billion dollars [8]. The failure or disruption of transportation networks, caused by accidents or nature disasters like hurricanes or snow storms, affects people's daily mobility [9, 10]. Flights might be cancelled and the travel time and travel distance might be increased due to the disruption [11]. Economic crisis, for example the global economic crisis [12] in 2008, highlights the need for a better

understanding of economic networks [13]. To make things worse, the failures in one infrastructure might propagate to other infrastructures due to the interdependency between different networks [14], for example, power networks and telecommunication networks [15].

How do we understand, characterize, quantify and improve the robustness of networks? Network theory is a powerful approach to investigate the robustness of networks. A brief history of the network theory is presented. In 1736, the great Swiss mathematician Leonhard Euler (1707-1783) solved the Königsberg bridges problem, as illustrated in Figure 1.1, which is regarded as the birth point of graph theory. In 1959, two Hungarian mathematicians Paul Erdös (1913 - 1996) and Alfréd Rényi (1921 - 1970) introduced random graphs [16] and established the random graph theory. Random graphs are extensively exploited in the field of complex networks to study the properties of graphs.

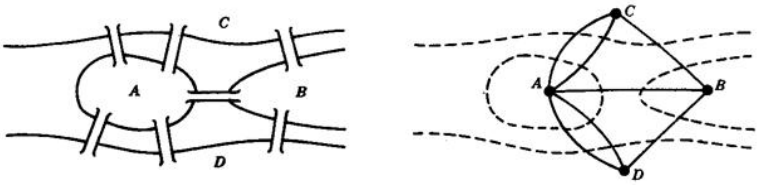
In 1967, the small-world phenomenon was observed in social networks. An experiment was performed by social psychologist Stanley Milgram (1933-1984) in the United States. The experiment aimed to figure out the number of social links between two randomly selected individuals in an acquaintance network. The experiment results showed that the number of social links on average is 5.5, which is known as "six degrees of separation".

In 1998, Watts and Strogatz [17] discovered small-world phenomenon in numerous real-world networks, including biological and technological networks. They proposed a model to generate small-world networks where (i) the average shortest path length between nodes is small, approximately in the order of the logarithm of the network size, and (ii) the clustering coefficient[1] is much higher than that in random graphs.

Real-world networks, such as World Wide Web, protein-protein interaction networks, e-mail networks exhibit properties that can not be captured by either random graphs proposed by Erdös and Rényi or small-world graphs proposed by Watts and Strogatz. In 1999, Barabási and Albert [18] unravelled the power-law degree distribution in networks like World Wide Web, known as scale-free networks. Barabási and Albert further argued that the scale free nature is rooted in network growth and

[1] Clustering coefficient is a graph metric reflecting the connection density among the neighbors of a node. The definition refers to Chapter 4

preferential attachment [18].



(a) Königsberg bridge (b) Euler's graphical representation

Figure 1.1: In the prussian city of Königsberg, there are four areas A, B, C, and D connected by seven bridges (Fig 1.1a). The problem is to devise a walk crossing each bridge once and only once. Euler simplified the problem by constructing a graph (Fig 1.1b) where each node represents an area and each link stands for each bridge and proved that such a walk is impossible.

The study of robustness, which is one of the early explored topics in complex networks, encounters two variants. The first one is the robustness of the topologies (maintenance of topological connectivity) of networks, called structural robustness, against failures of nodes or links. The second one is the robustness of the dynamical processes (maintenance of dynamical processes) running on networks, referred to as dynamical robustness. In 2000, Albert *et al.* [19] studied the structural robustness of complex networks against failures of nodes. The results show that scale-free networks display high tolerance to random failures while such networks are extremely vulnerable to targeted attacks. In 2000, a mathematical model, percolation model, which was first proposed by Broadbent *et al.* [20] in 1957, was employed to analytically study the structural robustness of networks [21, 22] followed by a series of studies [4, 23, 24]. The theory of generating functions [25] is applied to the percolation model in random graphs with arbitrary degree distribution [26].

The second ingredient of robustness needed to be accounted for is the dynamic process with emphasis on the interplay between the structure of a network and dynamics on that network. Real-world dynamics are, but not limit to, epidemic spreading in a population [27], flow distribution in power grids [28, 29], packets delivery in the Internet [30]. Characterizing the robustness of dynamical processes is in general complicate and difficult. Models, such as susceptible-infected-susceptible (SIS) and susceptible-infected-removed (SIR) [27], for epidemic spreading processes are proposed. The epidemic threshold, introduced by Kermack *et al.* [31] in 1927, is regarded as a robustness metric above which the epidemic persists

and below which epidemic dies out and the network is virus-free. Epidemic threshold, in the N-Intertwined Mean-Field Approximation (NIMFA), is shown to be inversely proportional to the spectral radius which is the largest eigenvalue of the adjacency matrix of a graph [27]. The synchronizability of a networks is characterized by the algebraic connectivity which is the second smallest eigenvalue of the Laplcacian matrix of a graph [32]. The successful applications of spectral metrics including another well-known example Google's PageRank [33], attract studies on the spectral domain [34] of networks.

The study on single networks is limited to anticipate the interaction between real-world networks. The study of robustness has lately switched to interdependent networks and focus on the understanding of the interconnection patterns between networks and how the interconnection influences the structural and dynamical robustness. In 2010, Buldyrev *et al.* [35] proposed a model of interdependent networks and showed that interdependent networks are subject to cascading failures. The discontinuous percolation transition behaves differently from the continuous phase transition in single networks. The dynamic epidemic spreading process in interdependent networks is characterized by the connection matrices of each coupled graph and the interconnection topology between the coupled networks [36].

1.1. RESEARCH QUESTIONS

The focus of this thesis is the robustness of complex networks including both theories and applications.

What is a robustness topology of a network against node or link failures? With a given network, how do we characterize and quantify its robustness? How do we add links into an existing network to maximally increase the robustness? Due to the interplay between network topologies and dynamic processes, how do we design a network that provides stable dynamic process, for example, the stable energy supply in power grids? How do we analyze the robustness of real-world metro networks?

Percolation models are employed to study the structural robustness of networks against node failures. How do we deal with robustness against link failures, for example the failure of transmission lines in power networks? How do we modify the existing methodologies for node failures to analyse the robustness against link failures? Shifting from topological domain, how do we understand and benefit from

the spectral domain of networks?

Since real-world networks interact with each other, how is the robustness of interdependent networks influenced by the interconnection pattern? How should the interconnection between networks be modelled? Do the dynamics in interdependent networks behave differently than the dynamics in single networks? When the dynamics in interdependent networks are separated into two regimes: (i) the interdependent network acts as a whole; (ii) the interdependent network operates as separated networks? This thesis dedicates to a better understanding of the above mentioned questions.

1.2. OUTLINE OF THIS THESIS

The thesis is organized in three parts. Part I presents robustness metrics and their applications in real-world networks. Part II focuses on the fundamentals of graph theory and part III investigates the robustness of interdependent networks.

1.2.1. PART I: ROBUSTNESS METRICS AND THEIR APPLICATIONS

In chapter 2, we investigate how to add a single link into an existing network such that the robustness is improved the most among all the possibilities. Based on the same principle, we study how to protect a link whose removal maximally decreases the robustness of a network.

In chapter 3, we discuss the application of a robustness metric, the effective graph resistance, in power systems. By adding single transmission lines, we improve the robustness of power grids against cascading failures.

In chapter 4, we analyse the robustness of 33 real-world metro networks by investigating ten theoretical and three numerical robustness metrics. We focus on which aspect of metros is captured by a robustness metric and thus provide insights for network planners on a robust design of metros.

1.2.2. PART II: FUNDAMENTALS OF GRAPH THEORY

Motived by the need to analyse robustness against link failures, we investigate a graph transformation, line graph. A line graph transforms links in the original graph to nodes in the line graph. In chapter 5, we study fundamental properties including

the degree distribution and the degree assortativity of the line graphs of complex networks.

Though the topological domain of a graph is widely studied, the spectral domain of a graph is less explored. Moving from structural properties of networks, Chapter 6 investigates the fundamental spectral properties of complex networks.

1.2.3. PART III: ROBUSTNESS OF INTERDEPENDENT NETWORKS

Chapters 7 and 8 start to focus on interconnection properties of real-world networks. Chapter 7 models the interconnection pattern for interdependent networks incorporating the locations of nodes. The robustness of interconnection patterns against node failures are evaluated and a new robustness metric that addresses the effect of a small fraction of failures, is proposed.

Chapter 8 studies the interdependent network consisting of two graphs with interconnections between them. The interconnections between the two graphs are represented by a weighted interconnection matrix B. We study the structural transition property for a regular interconnection matrix B (constant row and column sum).

In chapter 9, we summary the contributions of the thesis and discuss the future work in the field of the robustness of networks.

Part I: Robustness metrics and their applications

2

IMPROVING ROBUSTNESS VIA THE EFFECTIVE GRAPH RESISTANCE

2.1. INTRODUCTION

Several complex infrastructural networks are built to geographically distribute flows of critical resources for our society. Electrical networks, via power lines, and water/gas networks, via pipe lines, are representative examples. In the lines of these networks, opposition forces, governed by physical laws[1], resist the passage of electric current or water/gas molecules. It is shown that these physical characteristics of resistance in individual lines play a key role in the robustness of the network as a whole [37–39], e.g., network robustness under cascading failures [28].

This chapter studies the graph metric of *effective graph resistance* as a robustness measure of complex networks. The effective graph resistance can be measured in graphs, therefore, it is a robustness indicator for several real-world networks that can be modeled as graphs. Ellens *et al*. [37] show that the lower the effective graph resistance is, the more robust a network is. Adding a link reduces the effective graph resistance and thus improves the robustness of a network. This scenario is appli-

[1] The Ohm's law for electrical networks and the Poiseuille's law for water networks.

cable to infrastructural investments that shall increase system lifetime by installing single lines. On the other hand, removing a link increases the effective graph resistance. The robustness is improved by 'protecting' the link whose removal maximally increases the effective graph resistance. This scenario is applicable to cyber-physical targeted attacks of infrastructural lines. The challenge in both scenarios lies in the selection of a link, among all the possible ones, whose addition or removal maximally decreases or increases the effective graph resistance.

Earlier work studies the effective graph resistance in networks that are topologically changed. For example, Ghosh *et al.* [40] study the minimization of the effective graph resistance by allocating link weights in weighted graphs. Van Mieghem *et al.* [41] show the relation between the effective graph resistance and the linear degree correlation coefficient. Abbas *et al.* [39] reduce the effective graph resistance of a graph by adding links in a step-wise way. In contrast to the aforementioned approaches, this chapter focuses on the effective graph resistance as an indicator of robustness in complex networks when single links are added or removed.

The contributions of this chapter are the following: (i) Theorems that prove upper and lower bounds of the effective graph resistance. (ii) Optimization strategies that are experimentally evaluated under synthetic and real-world networks. These strategies maximize the decrease or the increase of effective graph resistance under link addition and removal respectively. (iii) A method and experimental results that topologically compare the optimal added or removed links according to effective graph resistance and algebraic connectivity. Therefore, this chapter provides a broad spectrum of theoretical and experimental findings on effective graph resistance as an indicator of robustness in synthetic and real-world networks.

This chapter is organized as follows: Section 2.2 defines the effective graph resistance and summarizes its properties. Section 2.3 derives bounds of the effective graph resistance under link addition and removal. The design and evaluation of the four strategies are illustrated in Section 2.4. The comparison between the optimization of the effective graph resistance and the algebraic connectivity is investigated in Section 2.5. Section 2.6 compares the optimization of the effective graph resistance with other approaches in related work. Section 2.7 concludes the chapter and outlines future work.

2.2. EFFECTIVE GRAPH RESISTANCE IN COMPLEX NETWORKS

Let $G(N,L)$ be an undirected graph with N nodes and L links. Adding or removing a link $e = i \sim j$ results in a graph $G+\{e\}$ or $G-\{e\}$. The adjacency matrix A of a graph G is an $N \times N$ symmetric matrix with elements a_{ij} that are either 1 or 0 depending on whether there is a link between nodes i and j or not. The Laplacian matrix Q of G is an $N \times N$ symmetric matrix $Q = \Delta - A$, where $\Delta = \text{diag}(d_i)$ is the $N \times N$ diagonal degree matrix with the elements $d_i = \sum_{j=1}^{N} a_{ij}$. The average degree in G is denoted as $E[D] = \frac{2L}{N}$. The Laplacian eigenvalues of Q are all real and non-negative [42]. The eigenvalues of Q are ordered as $0 = \mu_N \leq \mu_{N-1} \leq \ldots \leq \mu_1$. The second smallest eigenvalue $\mu_{N-1} = \alpha_G$ is coined by Fielder [43] as the algebraic connectivity. In this chapter, the effective graph resistance R_G is computed as follows [42]:

$$R_G = N \sum_{i=1}^{N-1} \frac{1}{\mu_i} \tag{2.1}$$

In order to compare the effective graph resistance R_G between networks with different size, the value of the effective graph resistance in Section 2.4 is normalized by dividing R_G with $\binom{N}{2}$.

The improvement of robustness via the effective graph resistance consists of two parts: adding an optimal link l_{R^+} that minimizes the effective graph resistance $R_{G+\{e\}}$ and protecting the link l_{R^-} whose removal maximizes the effective graph resistance $R_{G-\{e\}}$. The effective graph resistance strictly decreases if a link is added into a graph and strictly increases if a link is removed from a graph[2] [37, 44]. A strategy in this work refers to the addition of a single link $e = i \sim j$ according to a specific rule, with the aim to minimize the effective graph resistance of the graph $G+\{e\}$. The possible number of links that can be added is:

$$L_c = \binom{N}{2} - L \tag{2.2}$$

A strategy also selects a link to protect from all the possible links L whose removal maximally increases the effective graph resistance.

The comparison between the optimal link l_{R^+} for the effective graph resistance $R_{G+\{e\}}$ and the optimal link l_{α^+} for the algebraic connectivity $\alpha_{G+\{e\}}$ is based on two

[2] This is also confirmed by Section 2.3 based on interlacing [42].

computations. The two computations are also performed for the comparison between optimal links l_{R^-} and l_{α^-}.

The first computation calculates the probability that the two optimal links are the same link. From the definition (2.1) of the effective graph resistance R_G, the algebraic connectivity α_G can be written as $\alpha_G = \mu_{N-1} = \frac{1}{R_G/N - S}$, where $S = \sum_{k=1}^{N-2} \frac{1}{\mu_k}$. Based on the definition of S, an upper and lower bound of the algebraic connectivity in terms of the effective graph resistance is derived in the Appendix A. When S is negligibly low, the two optimal links for the algebraic connectivity α_G and for the effective graph resistance R_G are the same link with probability $\Pr[l_{R^+} = l_{\alpha^+}]$ for link addition and $\Pr[l_{R^-} = l_{\alpha^-}]$ for link removal.

The second computation concerns the distance between l_{R^+} and l_{α^+} when they are not the same link with probability $1-\Pr[l_{R^+} = l_{\alpha^+}]$. The distance between links in a graph G is measured by the hopcount in the corresponding line graph G^*. A line graph G^* of a graph G is a graph in which every node of G^* corresponds to a link in G and two nodes of G^* are adjacent if and only if the corresponding links in G have a node in common [42]. The graph G is referred to as the root graph of G^*. The links l_{R^+} and l_{α^+} in the root graph G are denoted as the nodes n_{R^+} and n_{α^+} in the line graph G^*. The hopcount $H(n_{R^+}, n_{\alpha^+})$ in G^* is the number of links in the shortest path between nodes n_{R^+} and n_{α^+}. The probability $\Pr[H(n_{R^+}, n_{\alpha^+}) = 0]$ equals to the probability $\Pr[l_{R^+} = l_{\alpha^+}]$. The hopcount $H(n_{R^+}, n_{\alpha^+}) = 1$ means that the link l_{R^+} and the link l_{α^+} share a common node.

Table 2.1 illustrates the mathematical symbols used in this chapter.

The complex networks in which this chapter focuses on include synthetic and real-world networks. Synthetic networks are as follows[3]:

Erdős-Rényi random graph [16] $G_p(N)$**:** This graph is generated from a set of N nodes by randomly assigning a link between each node pair with probability p. The probability p is also called the link density. When the link density p is higher than a critical threshold $p_c \approx \ln N/N$, the graph is connected [33].

Barabási-Albert power law graph [18]: This graph is generated by starting with m nodes. At every time step, a new node with m links is connected to the m existing nodes in the network. A new node connects to a node i in step t with probability $p = d_i/2L_t$, where d_i is the degree of node i and L_t is the total number of links at

[3] All these listed networks are converted to undirected and unweighted connected networks.

Table 2.1: An overview of the mathematical symbols

Symbol	Interpretation	Symbol	Interpretation
G	A graph	G^*	Line graph of a graph G
N	Number of nodes in a graph G	n_{R^+}	Node in line graph corresponding to l_{R^+}
L	Number of links in a graph G	n_{R^-}	Node in line graph corresponding to l_{R^-}
e	A link in a graph G	n_{α^+}	Node in line graph corresponding to l_{α^+}
A	Adjacency matrix	n_{α^-}	Node in line graph corresponding to l_{α^-}
a_{ij}	An element in the adjacency matrix A	$H(n_{R^+}, n_{\alpha^+})$	Hopcount between n_{R^+} and n_{α^+}
d_i	Degree of a node i	$H(n_{R^-}, n_{\alpha^-})$	Hopcount between n_{R^-} and n_{α^-}
Δ	Diagonal matrix with the nodal degrees	$\Delta\mu_i$	Increase or decrease of an eigenvalue μ_i
Q	Laplacian matrix	ρ	Diameter of a graph G
$E[D]$	Average degree	S_s	A strategy s
μ_i	Eigenvalue of the Laplacian matrix	y	Fiedler vector
α_G	Algebraic connectivity	R_{ij}	Effective resistance between nodes i and j
R_G	Effective graph resistance for a graph G	Q^{-1}	Moore-Penrose pseudoinverse of Q
C^*	Effective graph conductance	cc_i	Closeness centrality of a node i
$R_{G+\{e\}}$	Effective graph resistance for $G+\{e\}$	H_{ij}	Hopcounts from a node i to a node j
$R_{G-\{e\}}$	Effective graph resistance for $G-\{e\}$	$G_p(N)$	An Erdős-Rényi graph
l_{R^+}	Optimal link whose addition minimizes R_G	p	Link density
l_{R^-}	Optimal link whose removal maximizes R_G	l_{α^+}	Optimal link whose addition maximizes α_G
l_{α^-}	Optimal link whose removal minimizes α_G	$E[H]$	Average hopcount
L_c	Number of possible links for link addition	R_{D_s}	Relative difference of R_G
$E[R_{D_s}]$	Average of R_{D_s}		

time t.

Watts-Strogatz small-world graph [17]: This graph is generated from a ring lattice of N nodes and k links per node. Each link is rewired at random with probability p.

These graph models have characteristics found in real-world networks. For example, Erdős-Rényi graphs can model collaboration networks [45]. The world-wide web follows approximately a power law degree distribution [46]. Social networks are often connected as small world networks [17].

In this chapter the following real-world networks are considered:

Dutch Soccer Network [47]: A graph of the Dutch football in which players represent the nodes. Two nodes are connected if the corresponding two players have played together in a football match.

Coauthorship Network of Scientists [48]: Scientists are nodes and two scientists are considered connected if they are co-authors in one or more papers.

Protein-Protein Interaction Network[4]**:** The nodes are proteins and the links are pairwise protein-to-protein interactions.

Citation Network[5]**:** The nodes are scientific papers and the links between the nodes are citations.

Western States Power Grid Network [49]: The nodes represent transformers, substations and generators. The links represent high-voltage transmission lines.

Western European Railway Network [49]: The stations are the nodes and the links are lines between the stations.

2.3. THEORETICAL BOUNDS

Topological network changes influence various graph metrics such as the effective graph resistance and algebraic connectivity studied in this chapter. Upper and lower theoretical bounds measure the highest and lowest values that a graph metric can have after certain topological network changes. Therefore, bounds can be used to reason about robustness estimations under topological changes such as link addition or removal. Bounds provide valuable estimations in various application domains. For example, the upper and lower bounds of throughput instruct the design of a wireless network in which node connections follow mobility patterns

[4] http://www.pdb.org/pdb/home/home.do (Last accessed: Apr. 2014).

[5] http://vlado.fmf.uni-lj.si/pub/networks/data/ (Last accessed: Apr. 2014).

[50]. Another example is the estimation of interference by upper and lower bounds when nodes are clustered in Ad Hoc Networks [51].

2.3.1. LINK ADDITION

After adding a link e, resulting in a graph $G+\{e\}$, a lower bound of the effective graph resistance $R_{G+\{e\}}$ is derived in Theorem 1. An upper bound $R_{G+\{e\}} \leq R_G$ is obtained in the proof of Theorem 1 based on interlacing [42].

Theorem 1. *By adding a link e to a graph G, resulting in the graph $G+\{e\}$, the lower bound of the effective graph resistance $R_{G+\{e\}}$ is*

$$R_{G+\{e\}} \geq \frac{R_G}{1+\frac{\rho}{2}N} \tag{2.3}$$

where ρ is the diameter of G.

Proof. The sum of Laplacian eigenvalues equals [42]

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

After a link addition, graph G has $L+1$ links and it holds that $\sum_{j=1}^{N-1}(\mu_j+\Delta\mu_j) = 2(L+1)$. The increase of the eigenvalue $\Delta\mu_j$ satisfies $\sum_{j=1}^{N-1}\Delta\mu_j = 2(L+1)-\sum_{j=1}^{N-1}\mu_j = 2(L+1)-2L = 2$. Interlacing [42] $\mu_j \leq \mu_j+\Delta\mu_j \leq \mu_{j-1}$ shows that $\Delta\mu_j \geq 0$ for any j, so that $\Delta\mu_j \leq 2$. For positive real numbers q_1, q_2,..., q_n and real numbers a_1, a_2,..., a_n, it holds [42]

$$\min_{1\leq k\leq n}\frac{x_k}{a_k} \leq \frac{x_1+x_2+\ldots+x_n}{a_1+a_2+\ldots+a_n} \leq \max_{1\leq k\leq n}\frac{x_k}{a_k} \tag{2.4}$$

Let $x_j = \frac{1}{\mu_j+\Delta\mu_j}$ and $a_j = \frac{1}{\mu_j}$. Based on the definition (2.1) of the effective graph resistance, inequality (2.4) yields

$$\frac{1}{1+\max_{1\leq j\leq N-1}\frac{\Delta\mu_j}{\mu_j}} \leq \frac{\sum_{j=1}^{N-1}\frac{1}{\mu_j+\Delta\mu_j}}{\sum_{j=1}^{N-1}\frac{1}{\mu_j}} = \frac{R_{G+\{e\}}}{R_G} \leq 1$$

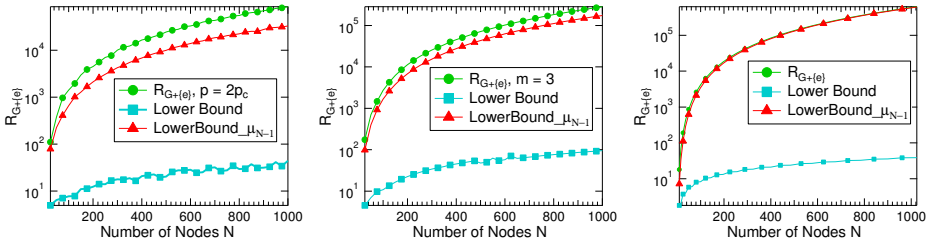
Furthermore, with $\max_{1\leq j\leq N-1}\frac{\Delta\mu_j}{\mu_j} \leq \frac{2}{\mu_{N-1}}$ and the lower bound [42] for the algebraic connectivity $\mu_{N-1} \geq \frac{4}{\rho N}$, the lower bound of (2.3) is derived. □

A consequence of the lower bound (2.3) is $\frac{R_{G+\{e_1\ldots e_m\}}}{R_G} \geq \left(1+\frac{m\rho}{2}N\right)^{-1}$ after m repeated link additions. In particular, a graph G can always be constructed by starting

from its minimum spanning tree and adding $L-N+1$ links. Given that the effective graph resistance $R_{MST} = \binom{N}{2} E[H_{MST}]$ for a minimum spanning tree [42], where H_{MST} is the hopcount in any minimum spanning tree, the lower bound of the effective graph resistance can be expressed as follows:

$$\begin{aligned} R_G &\geq \frac{R_{MST}}{1+\frac{\rho_{MST}}{2}N(L-N+1)} \\ &= \frac{\binom{N}{2}E[H_{MST}]}{1+\frac{\max H_{MST}}{2}N(L-N+1)} \end{aligned}$$

This bound may be valuable in sparse networks where L is not significantly larger than $N-1$.



(a) Erdős-Rényi graph (b) Barabási-Albert graph (c) Square lattice graph

Figure 2.1: Lower bounds of the effective graph resistance $R_{G+\{e\}}$.

Figure 2.1 shows the lower bound of the effective graph resistance $R_{G+\{e\}}$ from Theorem 1 in Erdős-Rényi, Barabási-Albert and square lattice[6] graphs. The lower bound is not tight, yet, a sharper lower bound can be derived by using the algebraic connectivity μ_{N-1} in the lower bound $\frac{R_G}{1+\frac{2}{\mu_{N-1}}}$. Figure 2.1 also shows the improved lower bound based upon the algebraic connectivity. This observation and the proof followed here suggest that the lower bound (2.3) can be improved with a sharper lower bound for the algebraic connectivity.

2.3.2. LINK REMOVAL

When a link e is removed from a graph, a lower bound of the effective graph resistance $R_{G-\{e\}}$ is derived in Theorem 2 and an upper bound in Theorem 3.

[6]The square lattice graph is a two-dimensional grid. Excluding the boundary nodes, the square lattice can be regarded as a regular graph with degree $d = 4$.

Theorem 2. *By removing a link e from a graph G, resulting in a reduced graph $G-\{e\}$, the lower bound of the effective graph resistance $R_{G-\{e\}}$ of the reduced graph $G-\{e\}$ is*

$$R_{G-\{e\}} \geq \frac{N(N-1)(N+1)}{2(L-1)} \tag{2.5}$$

where N is the number of nodes and L is the number of links of the original graph G.

Proof. Let $\Delta\mu_i$ defined as the amount of the decrease of an eigenvalue μ_i. The effective graph resistance $R_{G-\{e\}}$ of the reduced graph $G-\{e\}$ is

$$\begin{aligned} R_{G-\{e\}} &= N\sum_{i=1}^{N-1}\frac{1}{\mu_i-\Delta\mu_i} \\ &= N\left(\frac{1}{\mu_{N-1}-\Delta\mu_{N-1}}+\sum_{i=1}^{N-2}\frac{1}{\mu_i-\Delta\mu_i}\right) \end{aligned} \tag{2.6}$$

For positive real numbers a_1, a_2, …, a_n, the harmonic, geometric and arithmetic mean inequality [42] is

$$\frac{n}{\sum_{k=1}^{n}\frac{1}{a_k}} \leq \sqrt[n]{\prod_{k=1}^{n}a_k} \leq \frac{1}{n}\sum_{k=1}^{n}a_k \tag{2.7}$$

with equality only if all a_k are equal. Let a_1, a_2, …, a_n be equivalent to $\mu_{N-2}-\Delta\mu_{N-2}$, $\mu_{N-3}-\Delta\mu_{N-3}$, …, $\mu_1-\Delta\mu_1$ and $n=N-2$. Inequality (2.7) is expressed as follows:

$$\frac{N-2}{\sum_{i=1}^{N-2}\frac{1}{\mu_i-\Delta\mu_i}} \leq \frac{1}{N-2}\sum_{i=1}^{N-2}(\mu_i-\Delta\mu_i) \tag{2.8}$$

Taking the reciprocal and then multiplying $N-2$ on both sides of the inequality (2.8) yields

$$\begin{aligned} \sum_{i=1}^{N-2}\frac{1}{\mu_i-\Delta\mu_i} &\geq \frac{(N-2)^2}{\sum_{i=1}^{N-2}(\mu_i-\Delta\mu_i)} \\ &= \frac{(N-2)^2}{2(L-1)-(\mu_{N-1}-\Delta\mu_{N-1})} \end{aligned} \tag{2.9}$$

where the sum of eigenvalues satisfies $\sum_{i=1}^{N-1}(\mu_i-\Delta\mu_i)=2(L-1)$. Substituting the inequality (2.9) into (2.6) yields

$$R_{G-\{e\}} \geq N\left(\frac{1}{\mu_{N-1}-\Delta\mu_{N-1}}+\frac{(N-2)^2}{2(L-1)-(\mu_{N-1}-\Delta\mu_{N-1})}\right)$$

Since the function, for $x>0$,

$$f(x)=\frac{1}{x}+\frac{(N-2)^2}{2(L-1)-x}$$

has a unique minimum at the positive value $x = \frac{2(L-1)}{N-1}$, it holds that

$$f(x) \geq f(x_1) = \frac{(N-1)(N+1)}{2(L-1)}$$

which leads to the lower bound (2.5). □

Theorem 3. *By removing a link e, resulting in a graph $G-\{e\}$, the upper bound of the effective graph resistance $R_{G-\{e\}}$ of the reduced graph $G-\{e\}$ is*

$$\frac{R_{G-\{e\}}}{R_G} \leq \max_i \frac{\mu_i}{\mu_{i+1}}$$

where $i \in [1, N-2]$.

Proof. Let $x_k = \frac{1}{\mu_j - \Delta\mu_j}$ and $a_k = \frac{1}{\mu_k}$ in inequality (2.4), then

$$\frac{1}{1-\min_i(\frac{\Delta\mu_i}{\mu i})} \leq \frac{\sum_{i=1}^{N-1} \frac{1}{\mu_i - \Delta\mu_i}}{\sum_{i=1}^{N-1} \frac{1}{\mu_i}} \leq \frac{1}{1-\max_i(\frac{\Delta\mu_i}{\mu_i})} \tag{2.10}$$

After a link removal, the interlacing property [42] shows that,

$$\mu_{i+1} \leq \mu_i - \Delta\mu_i \leq \mu_i \tag{2.11}$$

where $i = 1, 2, \ldots, N-1$. Subtracting μ_i on both sides of (2.11) leads to

$$0 \leq \Delta\mu_i \leq \mu_i - \mu_{i+1} \tag{2.12}$$

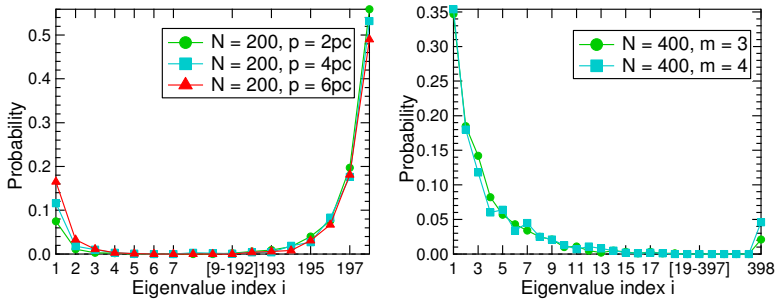
Substituting (2.12) into the right-hand side of (2.10) yields

$$\begin{aligned} \frac{1}{1-\max_i(\frac{\Delta\mu_i}{\mu_i})} &\leq \frac{1}{1-\max_i(\frac{\mu_i-\mu_{i+1}}{\mu_i})} \\ &= \frac{1}{1-(1-\min_i(\frac{\mu_{i+1}}{\mu_i}))} \\ &= \frac{1}{\min_i(\frac{\mu_{i+1}}{\mu_i})} = \max_i(\frac{\mu_i}{\mu_{i+1}}) \end{aligned}$$

Based on definition (2.1) of the effective graph resistance, we establish the theorem.

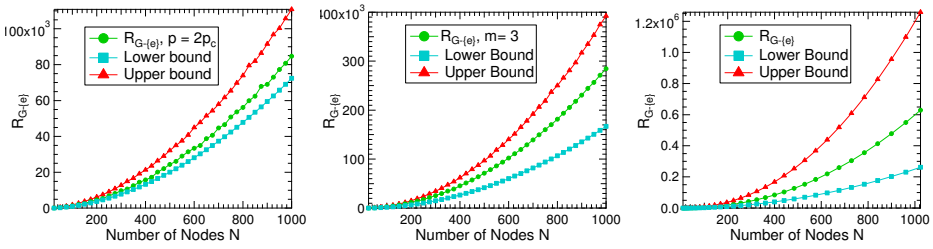
□

Figure 2.2 shows the probability that $\frac{\mu_i}{\mu_{i+1}}$ has a maximum at the index i within 10^3 instances of Erdős-Rényi and Barabási-Albert graphs, respectively. Figure 2.2a shows that $\frac{\mu_i}{\mu_{i+1}}$ has a maximum at $i = N-2$ with a probability higher than 0.5. Figure 2.2b shows that $\frac{\mu_i}{\mu_{i+1}}$ has a maximum at $i = 1$ with a probability 0.35. In both Figure 2.2a and 2.2b, the maximum of $\frac{\mu_i}{\mu_{i+1}}$ is attained within several highest and lowest values of the index i. Figure 2.3 shows the upper and lower bounds of the effective graph resistance $R_{G-\{e\}}$ from Theorem 2 and 3.



(a) Erdős-Rényi graph (b) Barabási-Albert graph

Figure 2.2: The probability that $\frac{\mu_i}{\mu_{i+1}}$ has a maximum at the index i in Erdős-Rényi and Barabási-Albert graphs.



(a) Erdős-Rényi graph (b) Barabási-Albert graph (c) Square lattice graph

Figure 2.3: Upper and lower bounds of the effective graph resistance $R_{G-\{e\}}$.

2.4. OPTIMIZATION OF THE EFFECTIVE GRAPH RESISTANCE

This section introduces four strategies for selecting a link whose addition minimizes the effective graph resistance and for protecting a link whose removal maximizes the effective graph resistance. The strategies are evaluated by comparing with the optimal effective graph resistance obtained by exhaustive search.

2.4.1. STRATEGIES FOR LINK ADDITION AND REMOVAL

In an exhaustive search, the optimal link l_{R^+} added between two nodes is discovered by checking all the possible links L_c. Similarly, the optimal link l_{R^-} is determined among all the possible links L.

An exhaustive search is computationally expensive as the number of nodes increases. More specifically, exhaustive search has a complexity order $O(N^5)$. This is computed by the computational order $\binom{N}{2} - L_c$ for checking all possible links multiplied by the order $O(N^3)$ for computing the pairwise effective resistance as illustrated in detail in Section 2.4.1. Strategies that determine the added or removed link based on topological and spectral properties of a network, provide a trade-off between a scalable computation and a high decrease or increase in the effective graph resistance. This section illustrates four strategies from which three of them are introduced in earlier work [52, 53], yet none of these strategies are evaluated for the effective graph resistance.

A strategy S_s, $s \in \{1, 2, 3, 4\}$, defines a link $e = i \sim j$, where e does not already exist under link addition and e already exists under link removal. The selection criteria of nodes i and j for each strategy are illustrated in the rest of this subsection. In this chapter, strategies S_1, S_2 are topological strategies and S_3, S_4 are spectral strategies.

SEMI-RANDOM - STRATEGY S_1

The node i has the minimum degree $\min(d_i)$ and node j is randomly chosen as $rand\{1, ..., L_c\}$.

The complexity of strategy S_1 is $O(N^2 - N + L_c + 1)$ computed as follows: (i) $O(N(N-1))$ is for counting the degrees of all the nodes. (ii) $O(L_c)$ is for finding the node i with minimum degree. (iii) $O(1)$ is for finding a random node.

DEGREE PRODUCT - STRATEGY S_2

The nodes i and j have the minimum[7] product of degrees $\min(d_i d_j)$. If there are multiple node pairs with the same minimum product of degrees, one of these pairs is randomly chosen.

The complexity of strategy S_2 is $O(N^2-N+2L_c)$ computed as follows: (i) $O(N(N-1))$ is for counting the degrees of all the nodes. (ii) $O(L_c)$ is for computing $d_i d_j$ for L_c

[7] Adding a link between nodes with the highest degree is evaluated as well. However, the performance is low and therefore this choice is not illustrated in this chapter.

Table 2.2: A summary of the strategies and the order of their computational complexity.

	Node i	Node j	Complexity Order
S_1	$\arg\min_i(d_i)$	$rand\{1,..., L_c \text{ or } L\}$	$O(N^2)$
S_2	$\arg\min_{i,j}(d_i d_j)$		$O(N^2)$
S_3	$\arg\max_{i,j}(\lvert y_i - y_j \rvert)$		$O(N^3)$
S_4	$\arg\max_{i,j}(R_{ij})$		$O(N^3)$

unconnected node pairs. (iii) $O(L_c)$ is for finding the minimum product $d_i d_j$.

FIEDLER VECTOR - STRATEGY S_3

The nodes i and j correspond to the i^{th} and j^{th} components of the Fiedler vector y that satisfy $\Delta y = \max(|y_i - y_j|)$, where $|y_i - y_j|$ is the absolute difference between the i^{th} and j^{th} components of the Fiedler vector y.

For strategy S_3, the complexity is $O(N^3 + 2L_c)$ computed as follows: (i) $O(N^3)$ is for computing the Fiedler vector y_i assuming the adoption of the QR algorithm [54] for computation. (ii) $O(L_c)$ is for computing $|y_i - y_j|$ for L_c unconnected node pairs. (iii) $O(L_c)$ is for finding the maximum of the difference $|y_i - y_j|$.

EFFECTIVE RESISTANCE - STRATEGY S_4

The nodes i and j have the highest effective resistance $\max(R_{ij})$. The pairwise effective resistance R_{ij} can be calculated as $R_{ij} = (\widehat{Q}^{-1})_{ii} + (\widehat{Q}^{-1})_{jj} - 2(\widehat{Q}^{-1})_{ij}$, where $\widehat{Q}^{-1}$ is the Moore-Penrose pseudoinverse [42] of Q.

For strategy S_4, the complexity is $O(N^3 + 4L_c)$ computed as follows: (i) $O(N^3)$ is for computing $\widehat{Q}^{-1}$. (ii) $O(3L_c)$ is for computing R_{ij} for L_c unconnected node pairs. (iii) $O(L_c)$ is for finding the maximum R_{ij}.

In case of link removals, L_c is replaced with L in all the four strategies. Table 2.2 summarizes all the strategies that add or remove a link $e = i \sim j$ and the order of their corresponding computational complexity.

The strategies illustrated in this chapter are indicative of a large number of other possible strategies. For example, two other strategies are tested:
S_5: The nodes i and j have the minimum product of closeness centrality $\min(cc_i cc_j)$. The closeness of a node i, $cc_i = \left[\sum_{j \neq i, j \in G} H_{ij}\right]^{-1}$, is computed as the inverse of the

sum of hopcounts H_{ij} from a node i to each node j.

S_6: The nodes i and j correspond to the i^{th} and j^{th} components of the principal eigenvector x_1 that have the maximum product $\max((x_1)_i(x_1)_j)$ of the eigenvector components. The principal eigenvector x_1 belongs to the highest eigenvalue of the adjacency matrix.

Strategy S_5 has higher complexity than S_1 and has approximately the same performance with S_1 for link addition. Strategy S_6 has the lowest performance under link addition and has approximately the same performance with S_2 for link removal. The rest of this chapter focuses on the four main strategies illustrated in this section.

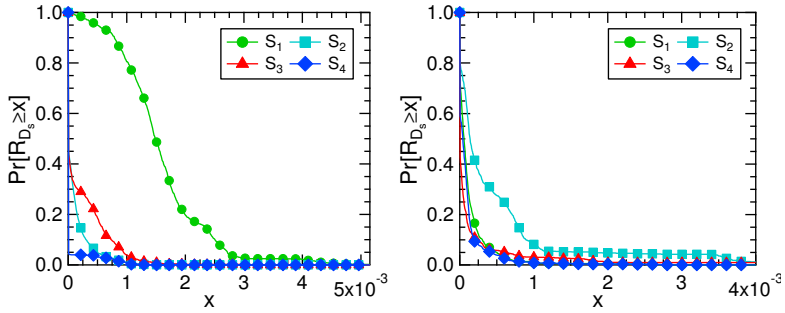
2.4.2. STRATEGY EVALUATION

The strategies are implemented and evaluated in MATLAB R2012b. First, the normalized optimal effective graph resistance $\overline{R^*}$ is obtained by applying exhaustive search. Second, the normalized effective graph resistance $\overline{R_{S_s}}$ is computed by adding or removing a link under each strategy $s \in \{1, 2, 3, 4\}$. Third, the absolute relative difference, $R_{D_s} = \left|\frac{\overline{R_{S_s}} - \overline{R^*}}{\overline{R^*}}\right|$ and the probability $\Pr[R_{D_s} \geqslant x]$, where $x \in [\min(R_{D_s}), \max(R_{D_s})]$, evaluate the performance of the four strategies. The lower the probability is, the closer R_{S_s} is to R^* and the more effective the strategy is. The average difference $E[R_{D_s}] = \int_0^\infty \Pr[R_{D_s} \geqslant x]\ dx$ computed by the area under the curve of the probability distribution, indicates the average performance of the strategies.

ERDŐS-RÉNYI RANDOM GRAPH

Figure 2.4 illustrates the performance of the four strategies in Erdős-Rényi random graphs. The figure is split into two subgraphs (a), (b), concerning link addition and removal. Figure 2.4a demonstrates that strategy S_4 is superior to all other strategies. Strategy S_2 outperforms strategy S_3 and strategy S_1 has the lowest performance. In Figure 2.4a, the average difference $E[R_{D_s}]$ for strategies S_1, S_2, S_3, S_4 is 2.99×10^{-3}, 0.24×10^{-3}, 0.36×10^{-3}, 0.04×10^{-3}.

Figure 2.4b shows that strategy S_4 is superior to S_3 and S_1. Compared to the second highest performance in Figure 2.4a, strategy S_2 has the lowest performance. The average difference $E[R_{D_s}]$ of strategies S_1, S_2, S_3, S_4 is 1.26×10^{-4}, 4.39×10^{-4}, 1.31×10^{-4}, 1.01×10^{-4}.

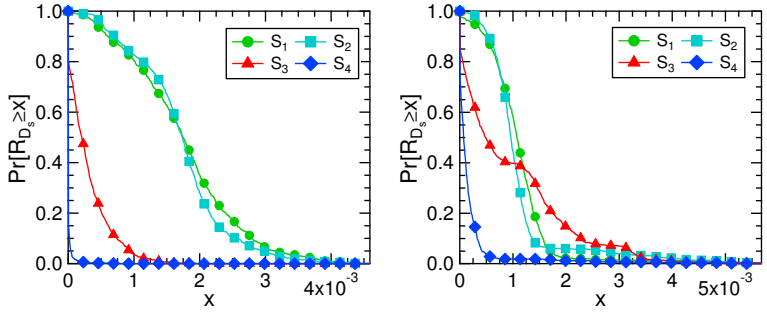


(a) Link Addition (b) Link Removal

Figure 2.4: $\Pr[R_{D_s} \geqslant x]$ for each strategy S_s, $s \in \{1,2,3,4\}$ in the Erdős-Rényi random graph with $N = 100$, $p = 2p_c$.

BARABÁSI-ALBERT POWER LAW GRAPH

Figure 2.5 illustrates the performance of the four strategies in Barabási-Albert power law graphs. Strategy S_4 achieves the highest performance in Figure 2.5a. Strategy S_3 outperforms strategies S_1 and S_2. The average difference $E[R_{D_s}]$ in Figure 2.5a for strategies S_1, S_2, S_3, S_4 is 1.74×10^{-3}, 1.69×10^{-3}, 0.29×10^{-3}, 0.01×10^{-3}.



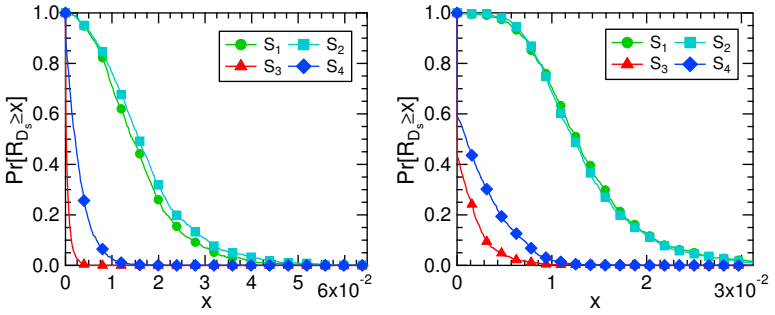
(a) Link Addition (b) Link Removal

Figure 2.5: $\Pr[R_{D_s} \geqslant x]$ for each strategy S_s, $s \in \{1,2,3,4\}$ in the Barabási-Albert power law graph with $N = 200$, $m = 3$.

Figure 2.5b shows strategy S_4 has the highest performance. The performance curve for S_3 crosses the curves for S_2 and S_1. Strategies S_2 and S_1 have comparable performance. The average difference $E[R_{D_s}]$ for strategy S_4 is 0.17×10^{-3}. For strategy S_3, the average difference $E[R_{D_s}]$ is 0.95×10^{-3} compared to 1.09×10^{-3} for strategies S_2, S_1, which indicates that strategy S_3 slightly outperforms S_2, S_1.

WATTS-STROGATZ SMALL-WORLD GRAPH

Figure 2.6 illustrates the performance of the four strategies in the Watts-Strogatz small-world graphs. In contrast to the results for Erdős-Rényi and Barabási-Albert, strategy S_3 outperforms strategy S_4 in both Figure 2.6a and 2.6b. Strategy S_1 is superior to S_2 in Figure 2.6a, while the opposite holds in Figure 2.6b.



(a) Link Addition (b) Link Removal

Figure 2.6: $\Pr[R_{D_s} \geqslant x]$ for strategy S_s, $s \in \{1,2,3,4\}$ in the Watts-Strogatz small world graph with $N = 100$, $k = 6$ and $p = 0.1$.

The average difference $E[R_{D_s}]$ for strategies S_1, S_2, S_3, S_4 in Figure 2.6a is 22.7×10^{-3}, 26.4×10^{-3}, 0.34×10^{-3}, 2.75×10^{-3}. These values in Figure 2.6b are 1.34×10^{-2}, 1.33×10^{-2}, 0.10×10^{-2}, 0.23×10^{-2}.

REAL-WORLD NETWORKS

Table 2.3 illustrates the performance of the four strategies in real-world networks. The table is ordered by the number of nodes in the network. The optimal added link by exhaustive search is not calculated because of the high computational complexity. Using Western States Power Grid Network as an example, the number of possible added links is 1.2×10^7. Therefore, the four strategies are evaluated by comparing the value of the effective graph resistance: the lower the effective graph resistance after link addition or the higher the effective graph resistance after link removal, the more effective the strategy.

For a given network, for example the *Dutch Soccer Network* in Table 2.3, the effective graph resistance of strategy S_3 is 0.1318 that is the lowest one compared to the effective graph resistance of S_1, S_2 and S_4. Strategy S_3 outperforms strategies S_1, S_2 and S_4. For all the listed networks except the *Western European Railway Network*

in Table 2.3, S_3 has the lowest effective graph resistance and outperforms the other three strategies. In contrast, the strategy S_4 outperforms strategy S_3 in *Western European Railway Network*. Strategy S_4 has the same performance as strategy S_3 in *Protein-Protein Interaction Network* and *Citation Network*.

Table 2.3: The effective graph resistance of the four strategies after link addition in real-world networks.

Name	N	L	R_{S_1}	R_{S_2}	R_{S_3}	R_{S_4}
Coauthorship	379	914	2.05	2.04	1.95	1.96
Protein	529	535	49.5	69.7	36.8	36.8
Dutch Soccer	685	10310	0.132	0.132	0.131	0.132
Citation	2678	10368	0.823	0.823	0.819	0.819
Power Grid	4941	6594	2.03	2.04	1.95	1.96
Railway	8710	11332	18.2	19.0	17.4	17.3

Table 2.4: The effective graph resistance of the four strategies after link removal in real-world networks.

Name	N	L	R_{S_1}	R_{S_2}	R_{S_3}	R_{S_4}
Coauthorship	379	914	2.08	2.07	2.21	∞
Protein	529	535	∞	∞	∞	∞
Dutch Soccer	685	10310	0.133	0.133	0.133	0.133
Citation	2678	10368	0.824	0.824	∞	∞
Power Grid	4941	6594	5.22	5.22	5.76	∞
Railway	8730	11332	19.0	19.0	19.4	∞

Table 2.4 shows the effective graph resistance of the four strategies under link removal. The infinite value of the effective graph resistance indicates that the removal of the selected link by a strategy disconnects the network. Strategy S_4 has the highest performance in all the listed networks. Strategy S_3 has comparable performance in *Protein-Protein Interaction Network*, *Dutch Soccer* and *Citation Network*.

PERFORMANCE OVERVIEW

Table 2.5 shows the ranking of the four strategies according to their performance. Strategy S_4 has the highest performance for both link addition and removal in Erdős-Rényi and Barabási-Albert graphs. In contrast, strategy S_3 has the highest

Table 2.5: The ranking of the four strategies according to their performance.

	Link Addition				Link Removal			
Rank / Network	1	2	3	4	1	2	3	4
Erdős-Rényi	S_4	S_2	S_3	S_1	S_4	S_1	S_3	S_2
Barabási-Albert	S_4	S_3	S_2	S_1	S_4	S_3	S_2	S_1
Watts-Strogatz	S_3	S_4	S_1	S_2	S_3	S_4	S_2	S_1
Real-world	S_3 or S_4		S_1 or S_2		S_3 or S_4		S_1 or S_2	

performance in Watts-Strogatz graphs under link addition and removal. Results are consistent with the larger graphs with number of nodes up to 400. In real world networks, either strategy S_3 or strategy S_4 has the highest performance for link addition and removal.

Despite the lower performance of strategies S_1 and S_2, their computational complexity is much lower compared to strategies S_3 and S_4. Therefore, the set of all strategies provides a trade-off between a low changing value of effective graph resistance and low computational complexity. Strategies S_1 and S_2 can be chosen when the computational resources are limited. Assuming that the computation of the optimal R^* is not an option for large networks, strategies S_3 and S_4 can be chosen under two scenarios: (i) In case of long term investments on infrastructural networks, such as railway networks, in which a link addition or removal is a costly operation and a strategy close to optimal R^* is a requirement. (ii) In case when the option of parallel computations, e.g. with MapReduce [55], is possible.

2.5. EFFECTIVE GRAPH RESISTANCE VS ALGEBRAIC CONNECTIVITY

The spectral expression of the effective graph resistance includes all the non-zero Laplacian eigenvalues, whereas the algebraic connectivity is one of the $N-1$ Laplacian eigenvalues. This section introduces a novel approach to compare the optimal links l_{R^+}, l_{α^+} and l_{R^-}, l_{α^-}. The comparison includes the probability that two optimal links are the same and the distance between the two optimal links in the corresponding line graph.

2.5.1. PROBABILITY OF THE SAME OPTIMAL LINK

Table 2.6 illustrates the probability $\Pr[l_{R^+} = l_{\alpha^+}]$ that l_{R^+} equals l_{α^+} in the 10^3 instances of Erdős-Rényi and Barabási-Albert graphs[8]. All the optimal links are obtained by exhaustive search. Table 2.6 illustrates that the maximum probability $\Pr[l_{R^+} = l_{\alpha^+}]$ obtained for Erdős-Rényi graph is 0.139 and for Barabási-Albert graph is 0.105. The optimal link for the algebraic connectivity is different from the optimal link for the effective graph resistance in most cases.

Table 2.6: The probability $Pr[l_{R^+} = l_{\alpha^+}]$ in Erdős-Rényi and Barabási-Albert graphs.

Erdős-Rényi	Probability	Barabási-Albert	Probability
$G_{2p_c}(50)$	0.139	$N = 100$, $m = 3$	0.034
$G_{2p_c}(100)$	0.102	$N = 100$, $m = 4$	0.105
$G_{2p_c}(200)$	0.074	$N = 200$, $m = 3$	0.013
$G_{4p_c}(200)$	0.068	$N = 200$, $m = 4$	0.066

Table 2.7 illustrates the probability $\Pr[l_{R^-} = l_{\alpha^-}]$ under link removal in the 10^3 instances of Erdős-Rényi and Bárabasi-Alber graphs. In contrast to the results in Table 2.6, the probability $\Pr[l_{R^-} = l_{\alpha^-}]$ is higher than 0.6 in Erdős-Rényi graph with link density $p = 2p_c$. However, when the link density p increases to $4p_c$, the probability $\Pr[l_{R^-} = l_{\alpha^-}]$ drops to approximately zero. One explanation is that the number of links in graph G increases with the increase of link density. The probability of choosing two links among all the possibilities decreases. The maximum probability $\Pr[l_{R^-} = l_{\alpha^-}]$ is 0.504 in Barabási-Albert graph. The decrease of the probability $\Pr[l_{R^-} = l_{\alpha^-}]$ with the increase of link density is also observed.

Table 2.7: The probability $Pr[l_{R^-} = l_{\alpha^-}]$ in Erdős-Rényi and Barabási-Albert graphs.

Erdős-Rényi	Probability	Barabási-Albert	Probability
$G_{2p_c}(50)$	0.677	$N = 100$, $m = 3$	0.504
$G_{2p_c}(100)$	0.665	$N = 100$, $m = 4$	0.208
$G_{2p_c}(200)$	0.613	$N = 200$, $m = 3$	0.460
$G_{4p_c}(200)$	0.002	$N = 200$, $m = 4$	0.113

[8] Results for the Watts-Strogatz small-world graphs are not included to keep the illustrations more compact. However, these results are available upon request

2.5.2. PROXIMITY OF OPTIMAL LINKS

This subsection illustrates how the distance between the optimal links l_{R^+} and l_{α^+} is computed when l_{R^+} is different from l_{α^+}. The hopcount in the corresponding line graph is proposed as a measure of the distance between the two optimal links l_{R^+} and l_{α^+}. Table 2.8 shows the average hopcount $E[H]$ between nodes n_{R^+} and n_{α^+} in the line graphs. In the line graphs of Erdős-Rényi graphs, the average hopcount between n_{R^+} and n_{α^+} approximates 1 that means the links l_{R^+} and l_{α^+} share a node in the original graph on average. The average hopcount between n_{R^+} and n_{α^+} in the line graphs of Barabási-Albert graphs approximates 2. From the definition of line graph, it can be derived that the end nodes of l_{R^+} and l_{α^+} are different but one of the end nodes of l_{R^+} is adjacent to one of the end nodes of l_{α^+}. Table 2.8 indicates that the optimal link for the algebraic connectivity is in a proximity of 1 or 2 hops to the optimal link for the effective graph resistance. This distance corresponds to 25% – 40% of the graph diameter.

Table 2.8: The average hopcount $E[H]$ between l_{R^+} and l_{α^+} in the Erdős-Rényi and Barabási-Albert graphs.

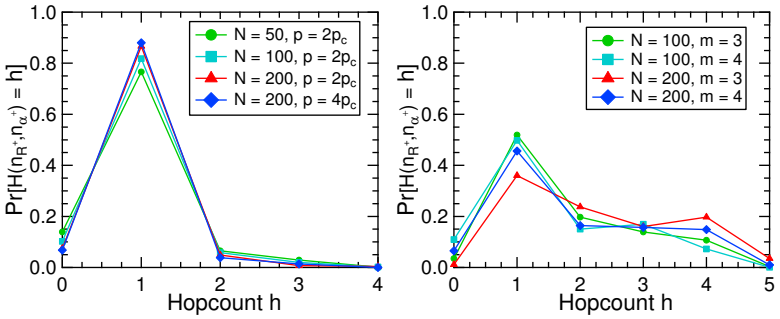
Erdős-Rényi	$E[H]$	Barabási-Albert	$E[H]$
$G_{2p_c}(50)$	0.987	$N = 100$, $m = 3$	1.759
$G_{2p_c}(100)$	1.002	$N = 100$, $m = 4$	1.636
$G_{2p_c}(200)$	1.001	$N = 200$, $m = 3$	2.285
$G_{4p_c}(200)$	0.998	$N = 200$, $m = 4$	2

As shown in Table 2.9, the average hopcount between n_{R^-} and n_{α^-} under link removal is lower than the average hopcount under link addition. For example, the $E[H]$ between n_{R^-} and n_{α^-} is 0.584 compared to 1.001 between n_{R^+} and n_{α^+} in Erdős-Rényi graph $G_{2p_c}(200)$. This observation is also confirmed by the fact that $\Pr[l_{R^-} = l_{\alpha^-}]$ is higher than $\Pr[l_{R^+} = l_{\alpha^+}]$.

Figure 2.7 illustrates the distribution of the hopcount $H(n_{R^+}, n_{\alpha^+})$ between the node n_{R^+} and n_{α^+} in the line graph of the Erdős-Rényi and Barabási-Albert graphs. In Figure 2.7a, the probability $\Pr[H(n_{R^+}, n_{\alpha^+})]$ is maximized for $H(n_{R^+}, n_{\alpha^+}) = 1$. The probability $\Pr[H(n_{R^+}, n_{\alpha^+}) > 1]$ converges to zero in 2 – 3 extra hops, especially for large N. In Figure 2.7b, the probability $\Pr[H(n_{R^+}, n_{\alpha^+})]$ is maximized for $H(n_{R^+}, n_{\alpha^+}) = 1$ and converges to zero for $H(n_{R^+}, n_{\alpha^+}) = 5$.

Table 2.9: The average hopcount $E[H]$ between l_{R^-} and l_{α^-} in the Erdős-Rényi and Barabási-Albert graphs.

Erdős-Rényi	$E[H]$	Barabási-Albert	$E[H]$
$G_{2p_c}(50)$	0.537	$N = 100$, $m = 3$	1.269
$G_{2p_c}(100)$	0.517	$N = 100$, $m = 4$	1.628
$G_{2p_c}(200)$	0.584	$N = 200$, $m = 3$	1.568
$G_{4p_c}(200)$	1.334	$N = 200$, $m = 4$	1.916



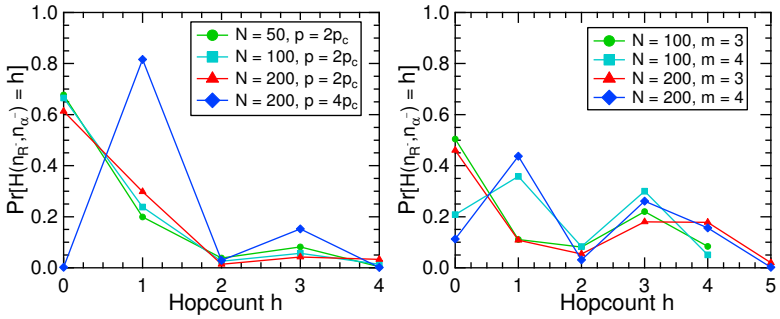
(a) Erdős-Rényi graph (b) Barabási-Albert graph

Figure 2.7: The distribution of the hopcount $H(n_{R^+}, n_{\alpha^+})$ in the line graph G^* between the node n_{R^+} and the node n_{α^+}.

Figure 2.8 illustrates the distribution of the hopcount $H(n_{R^-}, n_{\alpha^-})$ between the node n_{R^-} and n_{α^-} under link removal. In Figure 2.8a, the probability $\Pr[H(n_{R^-}, n_{\alpha^-})]$ is maximized for $H(n_{R^-}, n_{\alpha^-}) = 0$ with link density $p = 2p_c$. when link density p increases, the peak of the probability shifts from 0 to 1. The probability $\Pr[H(n_{R^-}, n_{\alpha^-}) > 1]$ converges to zero in $2-3$ extra hops. In Figure 2.8b, the peak of the probability $\Pr[H(n_{R^-}, n_{\alpha^-})]$ shifts from 0 to 1 as the average degree grows and the probability $\Pr[H(n_{R^-}, n_{\alpha^-}) > 1]$ converges to zero at $H(n_{R^-}, n_{\alpha^-}) = 5$.

2.6. COMPARISON WITH RELATED WORK

Network robustness is mostly studied under topological perturbations that usually concern (i) addition of nodes or links, (ii) removal of nodes or links, (iii) rewiring of links. These perturbations influence the spectral properties of networks. For example, Takamitsu *et al.* [56] study the influence of node removal on the second smallest Laplacain eigenvalue. Attilio *et al.* [57] focus on the largest eigenvalue un-



(a) Erdős-Rényi graph (b) Barabási-Albert graph

Figure 2.8: The distribution of the hopcount $H(n_{R^-}, n_{\alpha^-})$ in the line graph G^* between the node n_{R^-} and the node n_{α^-}.

der links perturbations. Van Mieghem *et al.* [53] study the spectral radius under link removal, whereas, Li *et al.* [58] investigate the spectral radius under node removal. In contrast to the spectral methodologies that consider a singe eigenvalue, the effective graph resistance studied in this chapter captures the information of all the eigenvalues and therefore it contains a broader range of spectral information about the network.

Various Internet protocols and applications transmit data packets via the shortest path between a source and destination. The effect of perturbations is studied by the changes of the shortest path length that is only one aspect influenced in the network. Holme *et al.* [59] introduce the average inverse length of shortest path as a measure of network robustness under perturbations. A higher shortest path length may result in slower information propagation in the network. This approach is limited to the evaluation of the changes on the shortest path length. However, effective graph resistance is a metric with a broader scope, e.g., power grid networks [60] in which power flows are transmitted via all possible paths besides the shortest path. In contrast to the measure of average shortest path length, the effective graph resistance is based on pairwise resistance that measures information of all the possible paths between a source and destination.

Furthermore, the study of topological perturbations in complex networks can be used for link prediction originated from information science. Link prediction refers to inferring added links in the near future or removed links from an observed

network [61]. Link prediction is applied in recommendation systems such as friendship recommendations between two strangers in social networks [62]. Algorithms based on structural nodal properties, such as the number of common neighbors [63] and an ensemble of all paths [62] are proposed for link prediction. Compared to structural properties, spectral characteristics of nodes provide different insights for link prediction, such as the Fiedler vector and effective resistance proposed in the optimization strategies of this chapter. Therefore, the link addition and removal strategies in this chapter can be potentially used in this application domain.

2.7. CHAPTER CONCLUSION

This chapter shows that adding or removing single links in theoretical and real-world complex networks has a measurable impact on network robustness. This chapter contributes theoretical and experimental findings that are applicable in real-world scenarios such as single-line installments in infrastructural networks or single-line protection against cyber-physical attacks. The upper and lower bounds introduced in this chapter can be used to support policy and decision makers to choose a line to install or protect given certain operational costs. Future work should study such trade-offs in specific application domain such as power grids. Moreover, when computational cost for finding optimal links to add or remove is prohibitive, the topological and spectral strategies studied in this chapter can still indicate links resulting in high robustness. This chapter also shows that if the optimal added or removed links for algebraic connectivity are known, then the respective links for effective graph resistance are different but in close proximity. Deriving analytically the optimal links of effective graph resistance given the optimal links of algebraic connectivity and vice versa, is a theoretical challenge to address in future work.

3

A NETWORK APPROACH FOR POWER GRID ROBUSTNESS

3.1. INTRODUCTION

The electrical power grid is crucial for economic prosperities of modern societies. Disruptions to electrical power grids paralyze the daily life and cause huge economical and social costs for these societies [64–66]. The strong dependency of other crucial infrastructures such as telecommunication, transportation and water supply on electrical power grids amplifies the severity of large scale blackouts [67]. The key importance of the power grid encourages further research into sustaining power system reliability and developing new approaches to evaluate and mitigate the risk of cascading blackouts.

Cascading failures are one of the main reasons for large scale blackouts [68]. Cascading failures are the consequence of the collective dynamics of a complex power grid. Large scale cascades are typically due to the propagation of a local failure into the global network [69]. Consequently, analyzing and mitigating cascading failures requires a system level approach. Recent advances in the field of network science [70] provide the promising potential of complex network theory to investigate the robustness of power grids at a system level. The robustness of power grids

in this chapter refers to their maintenance of function after cascading failures triggered by targeted attacks.

Analyzing and improving the network robustness includes two parts. The first goal is the proposal of a proper metric that characterizes the robustness of a specific class of networks [71]. A second goal is to propose efficient strategies on graph modification in order to increase the value of the proposed robustness metric. Consequently, an effective robustness metric that incorporates the essence of the power grids and effective strategies for graph modification are required to improve the robustness of power grids.

The effective graph resistance is a graph metric which characterizes the essence of electrical power grids such as power flow allocation according to Kirchhoff's laws. Researchers in [28] show that the effective graph resistance effectively captures the impact of cascading failures in a power grid. The lower the effective graph resistance is, the more robust a power grid is against cascading failures. Adding a link decreases the effective graph resistance [72]. This chapter focuses on enhancing the grid robustness against cascading failures by applying the effective graph resistance as a metric for network expansion.

Determining the right pair of nodes to connect in order to maximize the robustness is a challenge. Exhaustive search, i.e. checking all the possibilities, is computationally expensive. Compared to exhaustive search, this chapter proposes four strategies that provide a trade-off between a higher decrease of the effective graph resistance and a lower computational complexity.

Exhaustively evaluating the impact of each link addition on robustness reveals the occurrence of Braess's paradox in power grids. Braess's paradox, originally found in traffic networks [73], shows that adding a link can decrease the robustness of the network. Specific sub-structures that might result in Braess's paradox by adding an extra link are investigated. Simulation results indicate that the effective graph resistance effectively identifies a link whose addition increases the robustness while avoids the Braess's paradox. Moreover, most of the strategies highly increase the robustness at a low computational complexity.

This chapter is organized as follows: Section 3.2 introduces the model of cascading failures in power grids. Section 3.3 presents the computation of the effective graph resistance in power grids. Strategies to add a transmission line are illustrated

in Section 3.4. The experimental methodology is illustrated in Section 3.5 and the improvement of the grid robustness is evaluated in Section 3.6. Section 3.7 concludes the chapter.

3.2. MODEL OF CASCADING FAILURES IN POWER GRIDS

A power grid is a three-layered network consisting of generation, transmission and distribution parts. A graph can represent a power grid where nodes are generation, transmission, distribution buses and substations, and links are transmission lines. Additionally, links are weighted by the admittance (or impedance) values of the corresponding transmission lines.

Electrical power in a grid is distributed according to Kirchoff's laws. Accordingly, impedances, voltage levels at each individual power station, voltage phase differences between power stations and loads at terminal stations control the power flow in the grid. This chapter approximates the flow values in a grid by using a linear direct current (DC) flow equation that approximates the nonlinear alternative current (AC) power flow equation [74].

The maximum capacity C_l of a line l is defined as the maximum power flow that can be afforded by the line. As in [28], we assume that the maximum capacity of a transmission line is proportional to its initial load $L_l(0)$ as follows:

$$C_l = \alpha_l L_l(0) \tag{3.1}$$

where α_l is called the tolerance parameter of the line l.

In a power grid, transmission lines are protected by relays and circuit breakers. A relay of a transmission line measures the load of that line and compares the load with the maximum capacity C_l computed by equation (3.1). When the maximum capacity is violated, and this violation lasts long enough, the relay notifies a circuit breaker to trip the transmission line in order to prevent the line from permanent damage due to overloading. We assume a deterministic model for the line tripping mechanism. A circuit breaker trips at the moment the load of a transmission line exceeds its maximum capacity.

The failure of a transmission line changes the balance of the power flow distribution over the grid and causes a redistribution of the power flow over the network. This dynamic response of the system to this triggering event might overload other

transmission lines in the network. The protection mechanism trips these newly overloaded transmission lines and the power flow is again redistributed potentially resulting in new overloads. This cascading failure continues until no more transmission lines are overloaded.

3.3. EFFECTIVE GRAPH RESISTANCE IN POWER GRIDS

This section explains the complex network preliminaries, presents the effective graph resistance, and elaborates on how it is computed in electric power grids.

3.3.1. COMPLEX NETWORK PRELIMINARIES

The topology of complex networks can be represented by 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 by $L = |\mathcal{L}|$. Graphs with N nodes are completely described by an $N \times N$ adjacency matrix A, in which the element $a_{ij} = 1$ if there is a link between nodes i and j, otherwise $a_{ij} = 0$. In case of a weighted graph, the network is represented by the weighted adjacency matrix W where the element w_{ij} is a real number that characterizes a certain property of the link $i \sim j$. The weight can be distances in transportation networks, the delay in the Internet, the strength of the interaction in the brain networks, and so on.

The weighted Laplacian matrix $Q = \Delta - W$ of G is an $N \times N$ matrix, where $\Delta = \text{diag}(d_i)$ is the $N \times N$ diagonal degree matrix with the element $d_i = \sum_{j=1}^{N} w_{ij}$. The eigenvalues of Q are non-negative and at least one is zero [42]. Thus, the smallest eigenvalue of Q is zero. The eigenvalues of Q are ordered as $0 = \mu_N \leq \mu_{N-1} \leq \ldots \leq \mu_1$.

Graph metrics measure the structural and spectral properties of networks. The degree d_i of a node i specifies the number of connected neighbours to that node. The largest eigenvalue λ_1 (also called the spectral radius) of the adjacency matrix highly influences the dynamic processes on networks such as virus spreading and synchronization processes [75]. The eigenvector corresponding to the spectral radius is called principle eigenvector x_1 that characterizes the influence of link/node removal on spectral radius [53, 58]. The second smallest eigenvalue μ_{N-1} of the Laplacian matrix Q is coined by Fiedler [43] as the algebraic connectivity α_G. The corresponding eigenvector is called the Fiedler vector. The entries of the Fiedler vector provide a powerful heuristic for community detection and graph partition-

ing [76]. The strategies illustrated in Section 3.4 are based on these structural and spectral graph metrics.

3.3.2. EFFECTIVE GRAPH RESISTANCE IN POWER GRIDS

Effective resistance R_{ij} is the electrical resistance between nodes i and j computed by series and parallel manipulations when a graph is seen as an electrical circuit where each link in the graph has a unit resistance. According to the Ohm's law, the effective resistance is the potential difference between nodes i and j when a unit current is injected at node i and withdrawn at node j. The effective graph resistance R_G is the sum of the effective resistance over all pairs of nodes $R_G = \sum_{i=1}^{N} \sum_{j=i+1}^{N} R_{ij}$.

Computation of the effective graph resistance for a power grid necessitates the topology of the grid (i.e. interconnection of nodes) and reactance (or susceptance) values of the transmission lines in the grid. The weighted Laplacian matrix Q of a power grid reflects the interconnection of nodes by transmission lines. The weight w_{ij} corresponds to the susceptance (the inverse of reactance) value of the line $l = i \sim j$. The effective resistance R_{ij} between a pair of nodes is computed as [42]:

$$R_{ij} = \left(\hat{Q}^{-1}\right)_{ii} + \left(\hat{Q}^{-1}\right)_{jj} - 2\left(\hat{Q}^{-1}\right)_{ij} \tag{3.2}$$

where $\hat{Q}^{-1}$ is the Moore-Penrose pseudo-inverse of the Q.

In terms of eigenvalues of the weighted Laplacian matrix Q, the effective graph resistance can be written as [42]

$$R_G = N \sum_{i=1}^{N-1} \frac{1}{\mu_i} \tag{3.3}$$

where μ_i is the ith eigenvalue of Q. In this chapter, we use equation (3.3), which is computationally efficient, to compute the effective graph resistance.

3.4. STRATEGIES FOR ADDING A TRANSMISSION LINE

As a response to blackouts, additional transmission lines are placed aiming to increase the robustness of power grids. Determining the right pair of nodes to connect in order to maximize the robustness is the challenge. An exhaustive search, identifying the best pair of nodes to connect by checking all $L_c = \binom{N}{2} - L$ possibilities, is computationally expensive especially when the number of nodes increases.

Therefore, strategies that determine the transmission line to be added, provide a trade-off between a scalable computation and a high increase of the grid robustness.

The effective graph resistance is shown to be able to anticipate the robustness of power grids with respect to cascading failures [28]. This section investigates four strategies, studied in [77], for selecting a link whose addition potentially minimizes the effective graph resistance and accordingly maximizes the robustness. A strategy defines a link $l = i \sim j$ and the selection of nodes i and j for each strategy are illustrated in the rest of this section.

3.4.1. DEGREE PRODUCT

The nodes i and j have the minimum product of degrees $\min(d_i d_j)$. If there are multiple node pairs with the same minimum product of degrees, one of these pairs is randomly chosen.

The complexity for the strategy is $O(N^2 - N + 2L_c)$ computed as follows: (i) $O(N(N-1))$ is for counting the degrees of all the nodes. (ii) $O(2L_c)$ is for computing $d_i d_j$ for L_c unconnected node pairs and for finding $\min(d_i d_j)$.

3.4.2. PRINCIPLE EIGENVECTOR

The nodes i and j correspond to the i^{th} and j^{th} components of the principal eigenvector x_1 that have the maximum product $\max((x_1)_i(x_1)_j)$ of the principle eigenvector components. The principal eigenvector x_1 belongs to the largest eigenvalue of the weighted adjacency matrix W.

The complexity is $O(N^3 + 2L_c)$ computed as follows: (i) $O(N^3)$ is for computing the principle eigenvector x_1 assuming the adoption of the QR algorithm [54] for computation. (ii) $O(2L_c)$ is for computing $(x_1)_i(x_1)_j$ for L_c unconnected node pairs and for finding $\max((x_1)_i(x_1)_j)$.

3.4.3. FIEDLER VECTOR

The nodes i and j correspond to the i^{th} and j^{th} components of the Fiedler vector y that satisfy $\Delta y = \max(|y_i - y_j|)$, where $|y_i - y_j|$ is the absolute difference between the i^{th} and j^{th} components of the Fiedler vector [52].

The complexity is $O(N^3 + 2L_c)$ computed as follows: (i) $O(N^3)$ is for computing

Table 3.1: A summary of the strategies and the order of their computational complexity.

	Node i Node j	Complexity Order
DegProd	$\arg\min_{i,j}(d_i d_j)$	$O(N^2)$
PrinEigen	$\arg\max_{i,j}\left((x_1)_i (x_1)_j\right)$	$O(N^3)$
FiedlerVector	$\arg\max_{i,j}(\lvert y_i - y_j \rvert)$	$O(N^3)$
EffecResis	$\arg\max_{i,j}(R_{ij})$	$O(N^3)$
Exhaustive Search	$\arg\min_{i,j}(R_G)$	$O(N^5)$

the Fiedler vector y_i. (ii) $O(2L_c)$ is for computing $|y_i - y_j|$ for L_c unconnected node pairs and for finding $\max|y_i - y_j|$.

3.4.4. EFFECTIVE RESISTANCE

The nodes i and j have the highest effective resistance $\max(R_{ij})$, where R_{ij} is computed by equation (3.2).

The complexity is $O(N^3 + 4L_c)$ computed as follows: (i) $O(N^3)$ is for computing $\widehat{Q}^{-1}$. (ii) $O(4L_c)$ is for computing R_{ij} for L_c unconnected node pairs and for finding the maximum R_{ij}.

Table 3.1 summarizes all the strategies that identify a link $l = i \sim j$ and the order of their corresponding computational complexity. Table 3.1 also presents the complexity order of the exhaustive search in order to compare with the complexity of the four strategies. The complexity order $O(N^5)$ of the exhaustive search is computed by $O(N^2)$ for checking all the possibilities multiplied by $O(N^3)$ for computing the effective graph resistance after a link addition.

3.5. EXPERIMENTAL METHODOLOGY

The experimental method presented in this section evaluates the robustness of the improved power system against cascading failures triggered by deliberate attacks. This approach can be used to assess the performance of the effective graph resistance as a metric for link addition on improving the robustness of power grids. This section elaborates on attack strategies and the quantification of the grid ro-

bustness after cascading failures.

3.5.1. ATTACK STRATEGIES

This subsection designs attack strategies based on electrical node significance centrality and link betweenness centrality. The electrical node significance [66] is a flow-based measure for node centrality, specifically designed for power grids. The electrical node significance δ_i of a node i is defined as the total power P_i distributed by node i normalized by the total amount of power that is distributed in the entire grid:

$$\delta_i = \frac{P_i}{\sum_{j=1}^{N} P_j} \tag{3.4}$$

An attack based on δ_i refers to target the link incident to the node i that has the highest electrical node significance. Since node i has the number d_i of incident links, the link with the highest load is chosen.

The link betweenness centrality is a topological graph metric quantifying the centrality of a link in complex networks [33]. The betweenness centrality of a link is defined as the total number of the shortest paths that traverse the link l.

$$B_l = \sum_{i=1}^{N} \sum_{j=1}^{N} 1_{l \in \mathscr{P}(i,j)} \tag{3.5}$$

where $1_{\{x\}}$ is the indicator function: $1_{\{x\}} = 1$ if the condition $\{x\}$ is true, else $1_{\{x\}} = 0$, and $\mathscr{P}(i, j)$ is the shortest path between nodes i and j. An attack based on betweenness centrality targets the link with the highest betweenness centrality.

Placing an additional line according to different strategies (presented in Section 3.4) results in different improved power systems. In order to compare cascading damages of these improved systems, we always attack the same link identified by the node significance centrality or link betweenness centrality of the original power grid.

3.5.2. ROBUSTNESS EVALUATION

The robustness of power grids is evaluated by the criticality of the additional line and the damages after cascading failures triggered by targeted attacks. To assess the criticality of the newly added transmission line based on the effective graph resistance, we deploy an analogous approach as in [78]: the criticality of an added

line l in a graph G is determined by the relative decrease of the effective graph resistance ΔR_G^l that is caused by the addition of a link l:

$$\Delta R_G^l = \frac{R_G - R_{G+l}}{R_G} \tag{3.6}$$

where R_{G+l} is the effective graph resistance of the grid after adding a link l into G. Evaluation of equation (3.6) results in the theoretical robustness level of a power grid.

Initially, a transmission line identified by the four strategies and exhaustive search is added into the power grid. Then, the newly obtained grids are attacked and the cascading damages are quantified.

The damage caused by the cascade is quantified in terms of normalized served power demand DS: served power demand divided by the total power demand in the network. Computing the normalized served demand for an interval of tolerance parameters $[\alpha_{min}, \alpha_{max}]$ results in a robustness curve of a grid. The normalized area below the robustness curve is computed by a Riemann sum [33]:

$$r = \frac{\sum_{i=1}^{m+1} DS(\alpha_i)\Delta_\alpha}{\alpha_{max} - \alpha_{min}} \tag{3.7}$$

where the closed interval $[\alpha_{min}, \alpha_{max}]$ is equally partitioned by m points and the length of the resulting interval is $\Delta_\alpha = \frac{\alpha_{max} - \alpha_{min}}{m+1}$. $DS(\alpha_i)$ is the normalized served demand when the tolerance parameter of the network is $\alpha_i \in [\alpha_{min} + (i-1)\Delta_\alpha, \alpha_{min} + i\Delta_\alpha]$. Since the maximum value of DS is 1, $(\alpha_{max} - \alpha_{min})$ refers to the maximum possible area below the robustness curve ensuring that the value of r is between 0 and 1. Evaluation of equation (3.7) for the robustness curve results in the experimental robustness level of a power grid with respect to cascading failures.

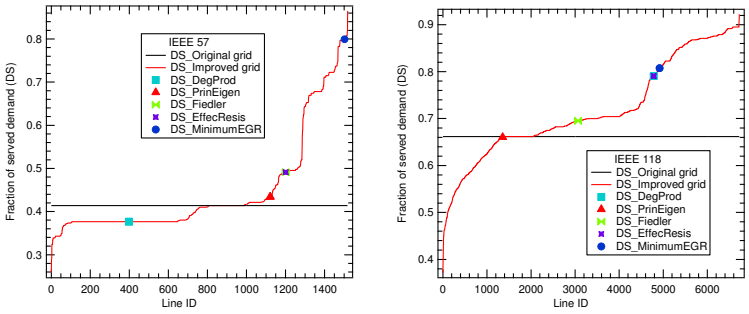
3.6. NUMERICAL ANALYSIS

This section investigates the effectiveness of the effective graph resistance as a metric for line addition, the impact of structures on the Braess's paradox, and the performance of the four strategies. First, the power grid is expanded by adding single links according to the minimization of the effective graph resistance, and the criteria of the four strategies. Then, the robustness of the improved power grid is assessed quantitatively under targeted attacks.

3.6.1. ASSESSING EFFECTIVENESS OF THE EFFECTIVE GRAPH RESISTANCE

Exhaustively adding all the possible links provides us all the possibly improved grids. Quantifying the cascading damages of all the improved grids under targeted attacks provides the benchmark for the evaluation of the effective graph resistance. The reactance value on each added line is assumed to be the average of all the existing transmission lines. The simulations are performed by MATCASC [79], a MATLAB based cascading failures analysis tool implementing the model in Section 3.2.

Figure 3.1 shows the performance of the effective graph resistance on identifying a critical link under a fixed tolerance parameter $\alpha = 2$ in IEEE 57 and 118 power test systems[1]. There are 1518 possible improved grids by adding a line to IEEE 57 and 6724 possible improved grids to IEEE 118. The original and improved power systems are attacked based on the node significance centrality computed by equation (3.4). In Figure 3.1, line ID is sorted in order of increasing normalized served demand DS. The horizontal line (i.e. the black line) is the served demand DS for the original power grid after cascading failures. The points on the red curve refer to the DS value of each improved grid that is obtained by adding one single line to the original network.



(a) IEEE 57 (b) IEEE 118

Figure 3.1: The performance of the effective graph resistance in IEEE 57, IEEE 118 power system with the tolerance parameter $\alpha = 2$.

The performance of the effective graph resistance as a metric for link addition and the performance of strategies are labelled in the Figure 3.1 with markers. The added line that minimizes the effective graph resistance increases the robustness

[1]IEEE power test systems: http://www.ee.washington.edu/research/pstca/

from 0.41 to 0.80 and improves the robustness by 95%. Compared to the possibly maximal increase 0.86 by a single link addition, the effective graph resistance achieves 93% accuracy in the IEEE 57 power system. Similarly in IEEE 118 power system, the added line that minimizes the effective graph resistance increases the robustness from 0.66 to 0.81. The effective graph resistance achieves 87% accuracy identifying the optimal line in the IEEE 118 power system.

In Figure 3.1, the curve above the horizontal line shows an increase of the robustness after a link addition, while the curve below the horizontal line presents a decrease of the robustness by adding a link. This counter-intuitive phenomenon is linked to Braess's paradox known for traffic networks, stating that adding extra capacity or links to a network occasionally reduces the overall performance of a network [73].

The simulation results in Figure 3.1 illustrate the effectiveness of the effective graph resistance to identify a critical link. The addition of the critical link improves the robustness of power grids regardless of the fact that the robustness can be decreased according to Braess's paradox. We further investigate more details on Braess's paradox in subsection 3.6.3.

3.6.2. ASSESSING THE EFFECTIVENESS OF STRATEGIES

To assess the effectiveness of the four strategies in Section 3.4, the IEEE 118 power system, consisting of 118 buses and 186 lines, is considered as a use case. For each line identified by each strategy, equation (3.6) is evaluated and its impact on the effective graph resistance is determined. Table 3.2 shows the lines to be added identified by strategies and their impact on the decrease of R_G.

Table 3.2: Added lines identified by the strategies and their impact on the decrease of R_G.

Strategy	line ID	$\Delta R_G^l(\%)$
DegProd	l_{87-117}	9.0
PrinEigen	l_{87-111}	4.2
Fiedler	$l_{111-117}$	11.3
EffectiveResis	l_{87-117}	9.0

In Table 3.2, the strategy based on the Fiedler vector selects the line connecting bus 111 and bus 117 and its addition causes 11.3% decrease of the effective graph resistance. Strategies based on the degree product and the effective resistance have an equal performance that decrease the effective graph resistance by 9%. The strategy based on the principle eigenvector decreases the effective graph resistance by 4.2%. Compared to other strategies, the strategy based on the Fiedler vector performs the best.

To validate the results from Table 3.2, the original and improved IEEE 118 power systems are attacked based on the electrical node significance and the link betweenness, and damages after cascading failures are quantified. The improved power system refers to the system after adding a transmission line identified by strategies in Section 3.4. Figures 3.2 and 3.3 show the robustness curves for improved power grids under an interval of tolerance parameters $[\alpha_{min}, \alpha_{max}]$ with $\Delta_\alpha = 0.05$, and highlight the improvement of the grid robustness. In order to quantify the performance of the four strategies in improving the grid robustness, the robustness value r in equation (3.7) for each robustness curve is shown in Table 3.3.

Figure 3.2 and Table 3.3 show the performance of the strategies in the IEEE 118 power grid under the attack based on the node significance. The strategy based on the Fiedler vector has a robustness value $r = 0.777$ which is an increase by 1.8% compared to the original grid robustness (i.e. 0.763). The strategy based on the degree product and on the effective resistance have an equal performance. These two strategies have the same robustness value $r = 0.769$ and increase the robustness by 0.8%. The strategy based on the principle eigenvector has the lowest performance and its robustness value is $r = 0.757$ that decreases the robustness by 0.8%.

Figure 3.3 and Table 3.3 present the performance of the strategies under the betweenness based attack. The strategy based on the Fiedler vector has the highest robustness value $r = 0.991$, which is an increase by 8.2% compared to the original grid robustness (i.e. 0.916). The strategy based on the degree product and on the effective resistance have an equal performance with the same robustness value $r = 0.949$. The robustness is increased by 3.6% compared to the original grid robustness. In contrast, the strategy based on the principle eigenvector with $r = 0.915$ slightly decreases the robustness by 0.1%. The performance order of the strategies shown in Figures 3.2 and 3.3 and Table 3.3 is in agreement with the theoretical results in Table

3.2.

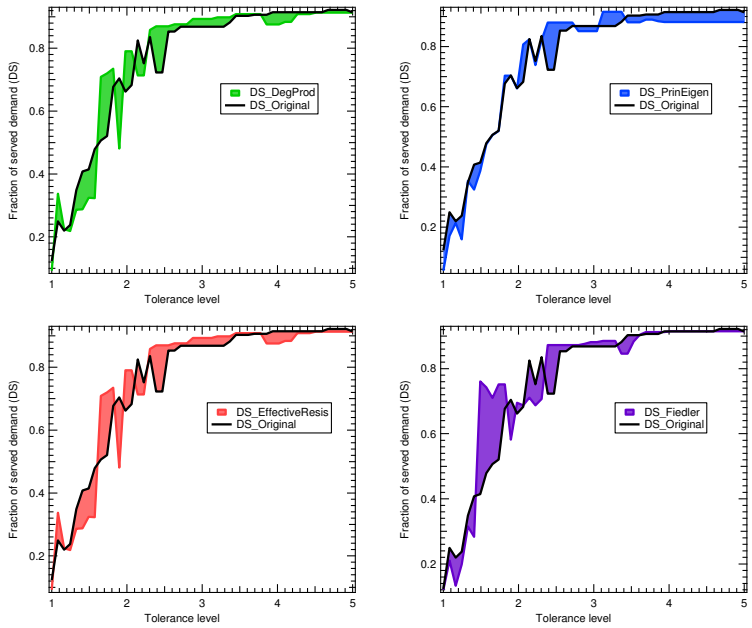


Figure 3.2: The performance of the four strategies in IEEE 118 power system under different tolerance parameters. The attack strategy is based on the node significance centrality.

Table 3.3: Critical lines identified by the four strategies and the robustness value r in IEEE 118 power system.

Strategy	line ID	r (Node Siginificance attack)	r (Betweenness attack)
DegProd	l_{87-117}	0.769	0.949
PrinEigen	l_{87-111}	0.757	0.915
Fiedler	$l_{111-117}$	0.777	0.991
EffectiveResis	l_{87-117}	0.769	0.949

When the computational cost for finding the optimal links to add is prohibitive, the strategy based on the Fiedler vector with the highest performance is preferable compared to other strategies. Assuming that computing the Fiedler vector for large grids is not an option, the strategy based on the degree product can be an alternative. The degree based strategy is more likely to be chosen than the strategy based

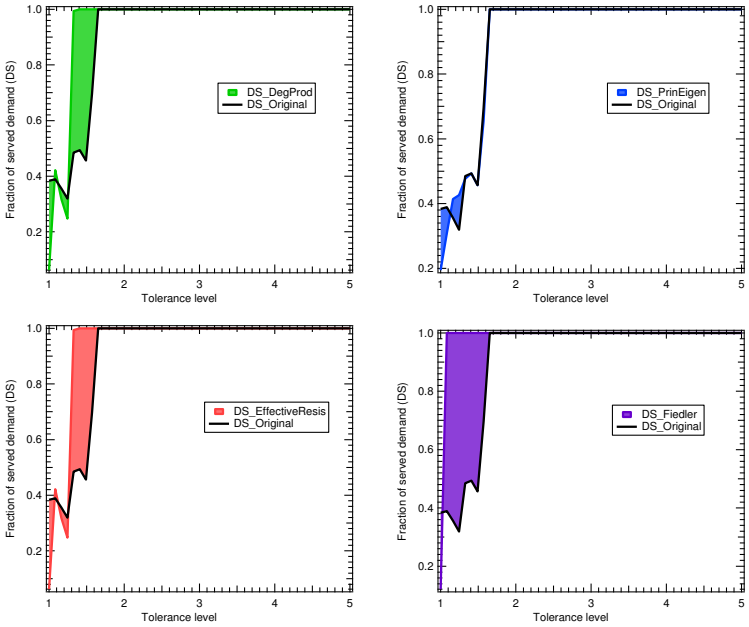


Figure 3.3: The performance of the four strategies in IEEE 118 power system under different tolerance parameters. The attack strategy is based on betweenness centrality.

on the effective resistance due to the fact that these two strategies have comparable performance, while the strategy based on the degree product has lower computational complexity.

3.6.3. ASSESSING THE IMPACT OF THE GRID TOPOLOGY ON BRAESS'S PARADOX

Braess's paradox in this chapter refers to the decrease of grid robustness by placing additional links. The relationship between the grid topology and the Braess's paradox in power grids is investigated.

The Wheatstone bridge graph (shown in Figure 3.4) refers to a graph consisting of four nodes, with four links creating a quadrilateral. A fifth link connects two opposite nodes in the quadrilateral, splitting the graph into two triangles [80]. We consider the subgraph with four nodes and four links as the Wheatstone subgraph and the fifth link as the Wheatstone link. Braess's paradox indicates that the construction of the Wheatstone bridge graph by adding the Wheatstone link occasionally

decreases the robustness of power grids. Let $P_{\text{Wheatstone}}$ represent the percentage of the Wheatstone links and P_{Paradox} be the percentage of the links, whose addition results in Braess's paradox. In order to investigate the impact of the Wheatstone bridge graph on Braess's paradox, the correlation between the percentages $P_{\text{Wheatstone}}$ and P_{Paradox} is quantified. The number of Wheatstone links is computed by the number of Wheatstone bridge subgraphs detected by FANMOD [81], a tool for fast network motif detection.

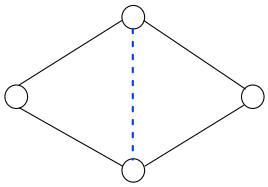
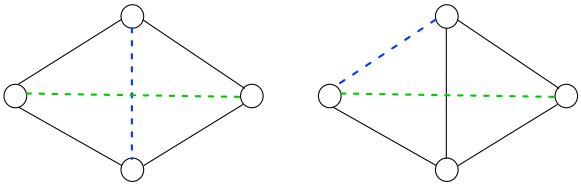


Figure 3.4: Wheatstone bridge graph

Figure 3.5 shows two types, Type I and Type II, of Wheatstone subgraphs from which a Wheatstone bridge graph is built by adding the Wheatstone link (the dashed line). For each subgraph, the number of the Wheatstone links is two times the total number of subgraphs of Type I and Type II. The percentage $P_{\text{Wheatstone}}$ of Wheatstone links in all the possible added links L_c is computed by $P_{\text{Wheatstone}} = \frac{2(N_{\text{TypeI}}+N_{\text{TypeII}})}{L_c}$, where N_{Typek} is the number of subgraphs of Type k. Table 3.4 shows the percentage $P_{\text{Wheatstone}}$ of Wheatstone links and the percentage P_{Paradox} in Figure 3.1. The correlation between $P_{\text{Wheatstone}}$ and P_{Paradox} is 0.96 suggesting the criticality of the Wheatstone bridge graph (see Figure 3.4) to the occurrence of Braess's paradox.

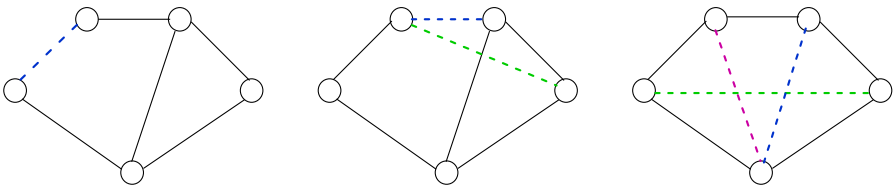
Besides the Wheatstone bridge graph that occasionally introduce Braess's paradox, we further investigate other subgraphs that may lead to the Braess's paradox. Figure 3.6 shows other three types, Type III to Type V, of subgraphs resulting in Braess's paradox when a single link is added. The dashed lines in Figure 3.6 are the possible links that cause the Braess's paradox. Table 3.5 shows the percentage $P_{\text{Wheatstone}}$ after including the number of links added into Type III, IV and V. The percentage $P_{\text{Wheatstone}}$ increases from 6.73% to 25.00% in IEEE 57 power system. An increase of the $P_{\text{Wheatstone}}$ from 4.53% to 15.44% is also observed in IEEE 118 and from 1.34% to 4.11% in IEEE 247 power system. Accordingly, the correlation be-



(a) Type I (b) Type II

Figure 3.5: Two types of subgraphs to build a Wheatstone bridge graph by adding the Wheatstone link. The dashed lines are the possible Wheatstone links.

tween $P_{\text{Wheatstone}}$ and P_{Paradox} increases to 0.971. The results indicate that the subgraphs from Type I to Type V provide an effective indication for the occurrence of the Braess's paradox in power grids.



(a) Type III (b) Type IV (c) Type V

Figure 3.6: Three types of subgraphs resulting in Braess's paradox by adding an extra link.

Table 3.4: The percentage $P_{\text{Wheatstone}}$ and P_{Paradox} in IEEE power systems

	IEEE57	IEEE118	IEEE247
L_c	1516	6717	30026
N_{TypeI}	0	20	30
N_{TypeII}	51	132	171
$P_{\text{Wheatstone}}(\%)$	6.73	4.53	1.34
P_{Paradox} (%)	53.16	20.67	4.57

Table 3.5: The percentage $P_{\text{Wheatstone}}$ and P_{Paradox} in IEEE power systems

	IEEE57	IEEE118	IEEE247
N_{TypeIII}	95	256	299
N_{TypeIV}	91	216	255
N_{TypeV}	0	15	8
$P_{\text{Wheatstone}}(\%)$	25.00	15.44	4.11

3.7. CHAPTER CONCLUSION

This chapter investigates the effective graph resistance as a metric for network expansions to improve the grid robustness against cascading failures. The effective graph resistance takes the multiple paths and their ability to accommodate power flows into account to quantify the robustness of power grids. The experimental verification on IEEE power systems demonstrates the effectiveness of the effective graph resistance to identify single links that improve the grid robustness against cascading failures. Additionally, when computational cost for finding optimal links is prohibitive, strategies that optimize the effective graph resistance can still identify an added link resulting in a higher level of robustness. Specifically, the strategy based on the Fiedler vector performs the best compared to other strategies and increases the robustness by 8.2% in IEEE 118 power system under the betweenness based attack, while reduces the computational complexity from $O(N^5)$ to $O(N^3)$.

The occurrence of Braess's paradox in power grids suggests that the robustness can be occasionally decreased by placing additional links. In particular, a badly designed power grid may cause enormous costs for new lines that actually reduce the grid robustness. The experimental results in this chapter provide insights in designing robust power grids while avoiding the Braess's paradox in power grids.

4

MULTI-CRITERIA ROBUSTNESS ANALYSIS OF METRO NETWORKS

4.1. INTRODUCTION

With constant urbanization [82], cities around the world are not only growing in number but they are also growing in size. As one of the main modes of urban transportation, public transit systems are integral to move people efficiently in cities [83–85]. Indeed, they provide myriads of benefits, from reducing traffic congestion to having a lesser impact on the environment, emitting fewer greenhouse-gases per capita than the conventional automobile [86, 87]. The future of public transportation is therefore bright. While increasing transit use is desirable, effort must be put into developing designs that are also resilient and robust. These subjects have gathered much interest in the scientific community in recent years, especially within the context of resilience to extreme events [88–90]. Resilience typically refers to the ability to return to a previous state after a disruption, while robustness tends to measure the amount of stress that can be absorbed before failure; Woods [91] inventoried four uses of the concept of resilience.

Traditionally, transit resilience and robustness have been associated largely with travel time reliability and variability [92]. It is still an important topic today from

quantifying variability itself [93, 94] or its cost [95], to using reliability and variability as a design criterion [96, 97]. Recently, the field of *Network Science* [98] has emerged as particularly fitted to measure the robustness of a system, notably by studying the impact of cascading failure [99–101]. Indeed, as physical networks, metros are composed of stations (nodes) and rail tracks (links), and they therefore possess measurable network properties [102, 103] that can be used to study their robustness [104–106]. Several works have also tried to combine information from both transit operation and network properties to gain insight into the robustness of transit networks [107–111].

In this work, our main objective is to analyse both theoretical and numerical robustness metrics for 33 worldwide metro systems within the realms of graph theory and network science. Metro, here, refers to heavy rail transit systems, whether underground, at grade, or overground. The freely available data from [112] was used[1].

To assess the robustness of metros, our main research approach is to subject metros to random failures and targeted attacks. Ten theoretical robustness metrics are investigated to anticipate the influence of failures and attacks in metro networks: (i) robustness indicator r^T, see [105], (ii) effective graph conductance C_G, see [42],(iii) reliability Rel_G, see [113], (iv) average efficiency $E[\frac{1}{H}]$, see [98], (v) clustering coefficient CC_G, see [98] (vi) algebraic connectivity μ_{N-1}, see [42] (vii) average degree $E[D]$, see [42] (viii) natural connectivity $\overline{\lambda}$, see [114] (ix) degree diversity κ, see [115] (x) meshedness coefficient M_G, see [116]. Moreover, the critical thresholds $f_{90\%}$ and f_c, see for instance [117], are obtained through simulations and categorize as numerical robustness metrics which provide the ground-truth for the robustness of metros under failures and attacks.

To evaluate whether the ten theoretical robustness metrics anticipate the metros robustness with respect to node failures, we investigate the Pearson correlations between theoretical and numerical robustness metrics. The strong correlations indicate that different robustness metrics quantify different aspects of robustness and highlight the multi-faced property of the robustness of metros. Finally, an overall robustness is provided by radar diagrams that incorporate all the ten robustness metrics.

The chapter is organized as follows. The definition and interpretation of theo-

[1] Available at http://csun.uic.edu/datasets.html, accessed July 8, 2016.

retical robustness metrics are studied in Section 4.2. Section 4.3 presents the simulation approach for numerical robustness metrics in 33 metro networks. The performance of the robustness metrics is assessed in Section 4.4. Section 4.5 concludes the chapter.

4.2. THEORETICAL ROBUSTNESS METRICS

This section elaborates on the ten theoretical robustness metrics and how these theoretical metrics relate to robustness of networks. A physical metro network can be represented by an undirected graph $G(N, L)$ consisting of N nodes and L links. The nodes are transfer stations and terminals, while the links are rail tracks that physically join stations. A graph G can be completely represented by an adjacency matrix A that is an $N \times N$ symmetric matrix with element $a_{ij} = 1$ if there is a connection between nodes i and j, otherwise $a_{ij} = 0$. The Laplacian matrix $Q = \Delta - A$ of G is an $N \times N$ matrix, where $\Delta = \text{diag}(d_i)$ is the $N \times N$ diagonal degree matrix with the elements $d_i = \sum_{j=1}^{N} a_{ij}$. The eigenvalues of Q are non-negative and at least one is zero [42]. The eigenvalues of Q are ordered as $0 = \mu_N \leq \mu_{N-1} \leq \ldots \leq \mu_1$. The degree $d_i = \sum_{j=1}^{N} a_{ij}$ of a node i is the number of connections to that node. The degree for the terminals is one.

4.2.1. THE ROBUSTNESS INDICATOR r^T

The *robustness indicator* r^T is suggested as a robustness metric for metro networks by Derrible and Kennedy [105]. It quantifies the robustness of a metro network in terms of the number of alternative paths in the network topology divided by the total number of stations in the system:

$$r^T = \frac{\mu - L^m}{N_S}$$

where N_S is the total number of stations (not limited to transfers and terminals), L^m is the number of multiple links between two nodes (e.g., overlapping lines), and μ is the cyclomatic number that calculates the total number of alternative paths in a graph; $\mu = L - N + P$, with L the number of links, N the number of nodes, and P the number of subgraphs. Transit networks are typically connected and, thus $P = 1$. The total number of stations, N_S in the denominator represents a likelihood of failure;

i.e., the larger the system, the more stations need to be maintained, and therefore the more likely a station may fail.

For this work, we do not consider any multiple edges[2]. Moreover, we also use the number of nodes N (i.e., transfer stations and terminals) in the denominator as opposed to the total number of stations N_S. Due to the sparsity of metro networks, i.e., $L < L_{max}$ with $L_{max} = \frac{N(N-1)}{2}$ obtained from the complete graph with N nodes, the robustness indicator in this chapter is modified as:

$$r^T = \frac{\ln(L-N+2)}{N} \tag{4.1}$$

where $\ln(L-N+2)$ is employed rather than $\ln(L-N+1)$ to avoid infinity for a tree graph with $L = N-1$. Essentially, r^T increases when alternative paths are offered to reach a destination, and it decreases in larger systems, which are arguably more difficult to upkeep. The normalized robustness indicator $\overline{r^T}$ is obtained dividing by $r^T = \frac{\ln(L_{max}-N+2)}{N}$ with $L_{max} = \frac{N(N-1)}{2}$.

4.2.2. THE EFFECTIVE GRAPH CONDUCTANCE C_G

The *effective graph resistance* R_G captures the robustness of a network by incorporating the number of parallel paths (i.e., redundancy) and the length of each path between each pair of nodes. The existence of parallel paths between two nodes in metro networks and a heterogeneous distribution of each path length result in a smaller effective graph resistance and potentially a higher robustness level.

The effective resistance R_{ij} [42] between a pair of nodes i and j is the potential difference between these nodes when a unit current is injected at node i and withdrawn at node j. The effective graph resistance R_G is the sum of R_{ij} over all pairs of nodes in the network. An efficient method for the computation of the effective graph resistance in terms of the eigenvalues is

$$R_G = N \sum_{i=1}^{N-1} \frac{1}{\mu_i}$$

where μ_i is the ith non-zero eigenvalue of the Laplacian matrix[3]. Properties of the effective graph resistance are given in [42]. The effective graph resistance is consid-

[2] Even when two stations are directly connected by multiple lines, we assign a value of 1 to the adjacency matrix. The definition is given in Section 4.3.1

[3] An $N \times N$ matrix representing the graph. The definition is given in Section 4.3.1.

ered as a robustness metric for complex networks [77], especially for power grids [28, 29]. In this chapter, we use a normalized version of the effective graph resistance, called the *effective graph conductance*, defined as

$$C_G = \frac{N-1}{R_G} \tag{4.2}$$

where C_G satisfies $0 \leq C_G \leq 1$. Here, a larger C_G indicates a higher level of robustness. The normalized C_G enables the comparison of network robustness among different cities with different metro size.

4.2.3. RELIABILITY

The *reliability* Rel_G of a network is the probability that the network is connected given the failure probabilities of its components. In this chapter, we model the reliability of each link specifically as opposed to the nodes. In the absence of actual reliability data (e.g., track maintenance and age), we use a constant value for the link reliability of 0.999 in accordance with values found in the literature [118] that includes, amongst others, vehicle breakdowns, power failures, and blockage. The reliability of a link is defined as one minus the failure probability, and the method assumes that the links have independent failure probabilities. This reliability measure is used often and in various contexts [119, 120], including in public transportation [121]. It essentially captures robustness by calculating the fraction of time every station is accessible from every other station. The downside of using the reliability is that it considers networks to be either fully operational or failed and does not provide any finer distinction. For further information, the reader is referred to [122].

4.2.4. AVERAGE EFFICIENCY $E[\frac{1}{H}]$

The hopcount H_{ij} is the number of links in the shortest path between node i and node j. The average hopcount $E[H]$ is defined as:

$$E[H] = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} H_{ij}$$

When a network is disconnected, the shortest paths between certain node pairs have infinite distance. To avoid an infinitely large metric under the scenario of a disconnected graph, the global average efficiency $E\left[\frac{1}{H}\right]$ is introduced by taking the

reciprocal hopcount between two nodes [98]:

$$E\left[\frac{1}{H}\right] = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{H_{ij}} \tag{4.3}$$

Assuming the transportation efficiency between two nodes is proportional to the reciprocal of their distance, the global efficiency quantifies the efficiency of transportation in a network on a global scale.

4.2.5. CLUSTERING COEFFICIENT CC_G

The clustering coefficient has become a standard in the network science literature to assess how the neighbors of a node are connected with one another. It was first introduced by [17]. The clustering coefficient of a node is defined as:

$$CC_i = \frac{2y_i}{d_i(d_i-1)}$$

where y_i is the number of links connecting neighbors of node i and d_i is the degree of node i. The clustering coefficient of a node i characterizes the connection density among the neighbors of node i. The maximum clustering coefficient is achieved in a complete graph where all the neighbors of a node are connected. In this work, we use the average clustering coefficient that is defined as the average of all individual clustering coefficients:

$$CC_G = \frac{1}{N} \sum_{i=1}^{N} CC_i \tag{4.4}$$

For a graph with N nodes, the clustering coefficient is bounded by

$$0 \leq CC_G \leq 1$$

where 0 is obtained in a tree and 1 is reached in a complete graph.

4.2.6. ALGEBRAIC CONNECTIVITY μ_{N-1}

The algebraic connectivity μ_{N-1} is the second smallest eigenvalue of the Laplacian matrix of a graph. When $\mu_{N-1} = 0$, the graph is disconnected whereas for $\mu_{N-1} > 0$ the graph is connected. It has been shown [42] that $\mu_{N-1} \leq \kappa_{\mathcal{N}}(G) \leq \kappa_{\mathcal{L}}(G)$ where $\kappa_{\mathcal{N}}(G)$ and $\kappa_{\mathcal{L}}(G)$ are node and link connectivity representing the minimum number of nodes and links whose removal disconnects the graph. Therefore, a high value

of the algebraic connectivity indicates a more robust network. In addition, it implies a strong synchrony in transport networks [123] and more difficulty to break down air transport networks [124] under random failures. Because the maximum algebraic connectivity for a graph with N nodes equals N, obtained for the complete graph, we normalize by dividing the algebraic connectivity by N. The normalized algebraic connectivity is denoted as $\overline{\mu_{N-1}}$.

4.2.7. AVERAGE DEGREE $E[D]$

For a graph with N nodes, the average degree can simply be written as:

$$E[D] = \frac{\sum_{i=1}^{N} d_i}{N} \tag{4.5}$$

where d_i is the degree of node i. Put simply, the average degree measures the number of average connections of a node. A network with a higher average degree can be thought of as more robust since it implies more connections (i.e., higher connectivity). We normalize the average degree dividing by the maximal degree, which is $N-1$, for a graph with N nodes. The normalized average degree is denoted as $\overline{E[D]}$.

4.2.8. NATURAL CONNECTIVITY $\overline{\lambda}$

The natural connectivity is defined as:

$$\overline{\lambda} = \ln\left[\frac{1}{N}\sum_{i=1}^{N} e^{\lambda_i}\right] \tag{4.6}$$

where λ_i denote the eigenvalues of the adjacency matrix of a graph. The natural connectivity characterizes the redundancy of alternative routes and is considered as a measure of structural robustness. The natural connectivity is a monotonical function of eigenvalue λ_i that is sensitive even to a single link failure [114]. Consequently, when link failures one by one, the natural connectivity is able to capture each failure, in contrast to, for instance, link connectivity that might be the same for certain link failures. The maximum natural connectivity for a graph with N nodes is obtained in the complete graph which is $N - \ln N$ as $N \to \infty$. In order to compare graphs with different sizes, we normalize the natural connectivity, denoted as $\overline{\lambda}^*$, dividing by the maximum natural connectivity $N - \ln N$.

4.2.9. DEGREE DIVERSITY κ

The degree diversity [115], also called the second-order average degree, is defined as:

$$\kappa = \frac{\sum_{i=1}^{N} d_i^2}{\sum_{i=1}^{N} d_i} \tag{4.7}$$

It has been shown that κ positively relates to the percolation threshold p_c [22] via $1 - p_c = \frac{1}{\kappa - 1}$ in the percolation model. The higher κ is, the more nodes need to be removed to disintegrate a network. In addition, the robustness of dynamic processes, e.g. epidemic spread, in a network relates to κ regarding the epidemic threshold [125], where below the epidemic threshold the network is safeguarded from long-term infection. As for homogeneous networks, such as regular graphs where each node has the same degree, the degree diversity tends to the average degree, $\kappa \to E[D]$. However, for scale-free networks with $N \to \infty$, the degree diversity tends to the infinity, $\kappa \to \infty$. In order to scale the value of the degree diversity in the interval $[0, 1]$, we take the inverse of the degree diversity.

4.2.10. MESHEDNESS COEFFICIENT M_G

The meshedness coefficient M_G is defined as:

$$M_G = \frac{L - N + 1}{2N - 5} \tag{4.8}$$

measuring the cycle structure in a planar graph by dividing the actual number of cycles by the potential number of cycles. It has notably been used to characterize the structural properties of urban street networks [116]. The difference between the meshedness coefficient M_G and the robustness indicator r^T lies in the denominator. The robustness indicator r^T considers the number of stations in the denominator, while M_G considers the maximal number of faces in a planar graph. The meshedness M_G satisfies $0 \leq M_G \leq 1$, where 0 is obtained in a tree graph with $L = N - 1$ and 1 is reached in the maximal planar graphs with $L = 3N - 6$.

4.3. NUMERICAL ROBUSTNESS METRICS

Numerical robustness metrics are obtained through simulations considering the robustness of 33 metro networks against random failures or deliberate attacks.

This approach can be used to evaluate the performance of different robustness metrics for metro networks under node failures/attacks. This section elaborates on the metro networks, attack strategies and determination of the critical thresholds.

4.3.1. METRO NETWORKS

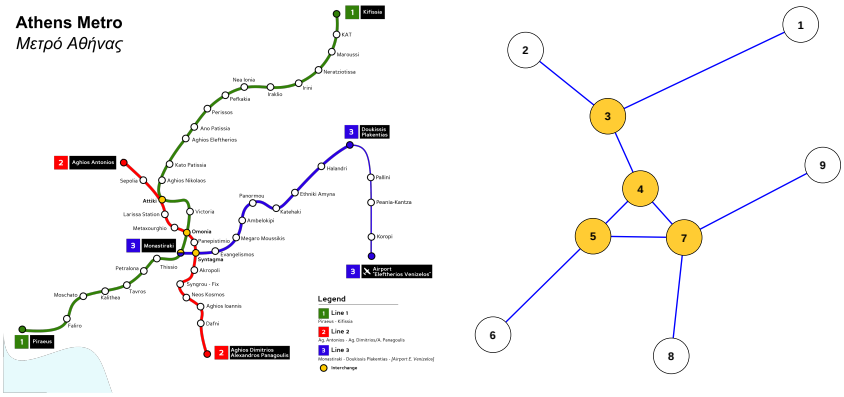
We define metros as urban rail transit systems with exclusive right-of-way whether they are underground, at grade or elevated. We represent a metro network by a graph, where nodes are transit stations and two nodes are connected if two transit stations are reachable. In this article, we look at 33 worldwide metro networks. Figure 4.1 exemplifies the graphical representation of a physical metro network. Figure 4.1(a) shows the map of the Athens metro network[4] and the graphical representation is shown in Figure 4.1(b). In Figure 4.1(b), stations 1 to 9 are respectively: Kifissia, Aghios Antonios, Attiki, Omonia, Monastiraki, Pireaus, Syntagma, Aghios Dimitrios, and Airport Eleftherios Venizelos. In this article, only the termini and transfer stations are taken into account, other stations that do not offer transfers or do not end lines are not considered as it was found preferable in [105, 112]. Moreover, they tend to bias the results by simply connecting with two adjacent stations. For more details on the methodology, see [105]. Note that the methodology presented here can be readily generalized for networks including non-transfer stations by considering weighted graphs instead of unweighed graphs, where the weights equal the number of non-transfer stations between two transfer stations plus one.

4.3.2. ATTACK STRATEGIES

To determine the robustness of metro networks, the response of metro networks to targeted attacks or random failures is investigated. This chapter considers two strategies for node removal: (i) random node removal and (ii) degree-based node removal.

- *Random removal:* The node to be removed is chosen at random from all the nodes in the network with equal probability.
- *Degree-based removal:* The node to be removed has the highest degree in the network. If multiple nodes have the highest degree, one node is chosen at ran-

[4] Adapted from http://commons.wikimedia.org/wiki/File:Athens_Metro.svg



(a) The map of the Athens metro network. (b) Graphical representation

Figure 4.1: Athens metro network

dom from all the highest-degree nodes with equal probability. In this chapter, nodes are removed progressively. We first remove the node with highest degree, and continue selecting and removing nodes in decreasing order of their degree.

4.3.3. CRITICAL THRESHOLDS

Critical thresholds relate to the fraction of nodes that have to be removed from the network, such that the size of the largest connected component of the remaining network is equal to a predetermined fraction of the size of the original network. Critical thresholds, which are also used in the percolation model [21, 126], characterize the robustness of interconnection patterns with respect to the removal/failure of network nodes.

After a node is removed, the size of the largest connected component of the remaining network is determined. Measuring the size of the largest connected component for an interval of removed nodes $[1, N]$ results in a robustness curve. From the robustness curve, we then determine the critical thresholds $f_{90\%}$ and f_c. The critical threshold $f_{90\%}$ is the first point at which the size of the largest connected component is less than 90% of the original network size. When determining the $f_{90\%}$ for random node removal, the size of the largest connected component is the average of 1000 simulation runs. Similarly, the critical threshold f_c is the first point at which

the size of the largest connected component is one (i.e., the network is completely disintegrated). Figure 4.2 exemplifies the determination of the critical thresholds from the robustness curve in *Tokyo* metro network with 62 nodes. Computing the size of the largest connected component for removed nodes from 1 to 62 results in a robustness curve. The size of the largest connected component is 56.77 after randomly removing 4 nodes. After removing 5 nodes, the size becomes 55.48 which is smaller than 90% × 62 = 55.8, i.e., 90% of the size of the network. Therefore, the critical threshold $f_{90\%}$ is determined as $\frac{5}{62}$. The threshold f_c is determined in a similar way. The critical thresholds are regarded as the experimental robustness level of metro networks with respect to node failures.

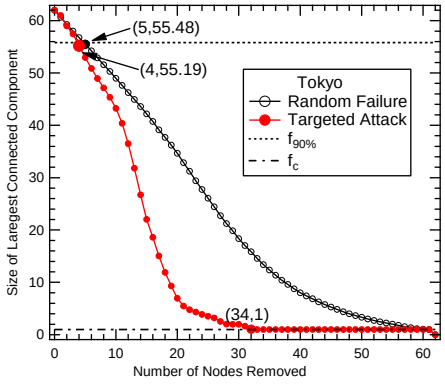


Figure 4.2: The robustness curve for the Tokyo metro network.

In this chapter, we first consider the threshold $f_{90\%}$, the fraction of nodes that have to be removed such that the remaining network has a largest connected component that contains 90% of the original network. For the node removal process, we simulate both random failures and targeted attacks. In the case of *random failures*, the nodes are removed by random selection, while for *targeted attacks*, the nodes are removed progressively based on their degrees (i.e., stations with many connections are removed first).

For the targeted attacks and random failures, we also consider the *critical threshold* f_c defined as the fraction of nodes to be removed such that the largest component is reduced to a size of one node (i.e., the network is completely disintegrated). As opposed to the theoretical metrics discussed in Section 4.2, the critical thresholds $f_{90\%}$ and f_c are obtained through simulations.

4.4. METRIC ANALYSIS FOR METRO NETWORKS

In this section, we study the robustness metrics for the 33 metro networks. Firstly, the ten theoretical robustness metrics are computed for the 33 metro networks. Secondly, the critical thresholds of metro networks under random failures and targeted attacks are determined by simulations. Thirdly, the relationship between the theoretical robustness metrics and numerical robustness metrics is studied. Finally, the overall performance of all the robustness metrics for the 33 metros is investigated.

4.4.1. EFFECTIVENESS OF ROBUSTNESS METRICS

Table 4.1 shows the values of the ten robustness metrics (from column 4 to column 13) computed using equations (4.1) to (4.8) and the four numerical robustness metrics (from column 14 to column 17) using the algorithms described in Section 4.3.3 for the 33 metro networks.

According to the rank of the robustness indicator $\overline{r^T}$, the most robust network is *Tokyo* with $\overline{r^T} = 0.512$, followed by *Madrid* and *Paris* with $\overline{r^T} = 0.5$ and 0.488, respectively. Moreover, *Seoul*, *Moscow* and *MexicoCity* also have a relatively high robustness level. Clearly, the robustness indicator $\overline{r^T}$ favors larger networks that have developed many alternative paths between any pairs of nodes. At the same time, $\overline{r^T}$ discredits networks that have a high number of nodes while having few alternative paths. This is particularly exemplified by the case of *New York*. Due to the topography of the region, the *New York* metro lines run mostly North-South from the Bronx to Lower Manhattan and East-West in Queens and Brooklyn. The lines therefore seldom intersect as opposed to the case of the *Seoul* metro for instance.

According to the effective graph conductance C_G, *Rome* with $C_G = 0.25$ has the highest robustness level, followed by *Cairo* and *Marseille* both with $C_G = 0.17$. The effective graph conductance accounts for the number of alternative paths, but it emphasizes on the length of each alternative path. For instance, for smaller networks without cycles (e.g., star graph), the effective graph conductance increases due to the lower average path length between two stations. The topologies in Figure 4.3a and Figure 4.3b are particular examples. In this case, a higher effective graph conductance indicates a lower number of transfer hops between two transit stations. At the same time, effective graph conductance favors networks with

the smallest length of the shortest paths. Taking Figure 4.3c (*Montreal*) and Figure 4.3d[5] (*Toronto*) as examples, the difference between the topologies is that station 1 connects to 10 and then connects to station 3 in *Toronto*, while stations 1 and 10 separately connect to stations 2 and 3 in *Montreal*. The total length of shortest paths from station 1 to the rest of the stations is higher in *Toronto* than in *Montreal*. Compared to *Toronto*, the higher effective graph conductance in *Montreal* indicates that the effective graph conductance favors the star-like topology with a smaller average shortest path length.

The reliability Rel_G indicates, just as the effective graph conductance does, that *Rome* is the most robust network with $Rel_G = 0.996$. After this, the most robust networks according to their reliability are *Bucharest*, *Cairo* and *Marseille*, each with $Rel_G = 0.995$. Of these three, *Cairo* and *Marseille* are also in second place according to the effective graph conductance. The reliability is sensitive to "bridges" in the network. In this work, a "bridge" is an link that if removed disconnects the network. They are of importance for the reliability because these edges must always be operational if the network is to remain a single connected component. Using this definition, we see that *Rome* has four bridges and the three networks following have five. The network with the lowest reliability is *London*. This is also the network with the most nodes and with the most bridges. Metro networks are often scale-free [105], which means that larger networks have more degree one nodes (the links to these nodes are always bridges). Therefore, it makes sense that the largest network has the highest amount of bridges and is the least reliable. Of course with different link reliabilities this line of reasoning would not hold any more.

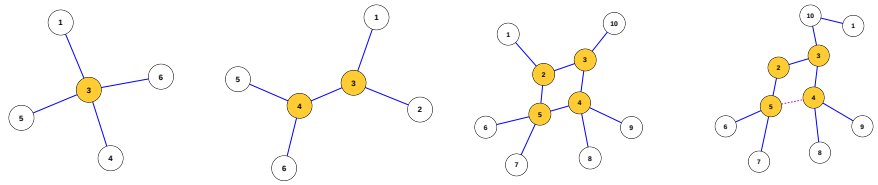
According to the rank of $\overline{r^T}$, $\frac{1}{\kappa}$ and M_G, *Tokyo* is the most robust metro network compared to other 32 metros. Meanwhile, according to C_G, Rel_G, $E[\frac{1}{H}]$, $\overline{\mu_{N-1}}$, $\overline{E[D]}$ and $\overline{\lambda}^*$, *Rome* is the most robust metro. *Barcelona* is considered as a robust network by the clustering coefficient CC_G. *Madrid* has a relatively high robustness level favoured by $\overline{r^T}$ and M_G. *Tokyo* and *Paris* are considered as robust networks by CC_G and $\frac{1}{\kappa}$, respectively. *Cario* and *Marseille* have a relatively high robustness level regarding the second highest value of metrics C_G, Rel_G, $E[\frac{1}{H}]$, $\overline{\mu_{N-1}}$, $\overline{E[D]}$ and $\overline{\lambda}^*$. The differences in these results suggest that robustness is a multi-faceted no-

[5] In order to compare the topology of *Montreal* and *Toronto*, a link between stations 4 and 5 is added into *Toronto* and the effective graph conductance is 0.099.

Table 4.1: Robustness metrics in 33 metro networks

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Metros	**N**	**L**	r^T	C_G	Rel_G	$E[\frac{1}{H}]$	CC_G	$\overline{\mu_{N-1}}$	$\overline{E[D]}$	$\overline{\lambda}^*$	$\frac{1}{\kappa}$	M_G	$f_{90\%\text{-Degree}}$	$f_{90\%\text{-Random}}$	$f_{c\text{-Degree}}$	$f_{c\text{-Random}}$	Area
Athens	9	9	0.206	0.11	0.994	0.54	0.09	0.031	0.25	0.14	0.38	0.08	0.11	0.11	0.44	0.89	0.63
Barcelona	29	42	0.456	0.03	0.987	0.37	0.17	0.006	0.1	0.06	0.26	0.26	0.07	0.03	0.62	0.97	0.81
Berlin	32	43	0.417	0.03	0.986	0.36	0.08	0.005	0.09	0.05	0.28	0.2	0.09	0.06	0.56	0.97	0.45
Boston	21	22	0.209	0.03	0.984	0.37	0.03	0.005	0.1	0.06	0.34	0.05	0.05	0.05	0.43	0.95	0.17
Brussels	9	9	0.206	0.11	0.994	0.55	0.09	0.034	0.25	0.14	0.38	0.08	0.11	0.11	0.44	0.89	0.64
Bucharest	11	12	0.287	0.1	0.995	0.52	0.06	0.036	0.22	0.12	0.35	0.12	0.09	0.09	0.45	0.91	0.63
BuenosAires	12	13	0.273	0.09	0.992	0.52	0.08	0.03	0.2	0.12	0.28	0.11	0.08	0.08	0.33	0.92	0.68
Cairo	6	5	0	0.17	0.995	0.62	0	0.073	0.33	0.18	0.45	0	0.17	0.17	0.33	0.83	0.79
Chicago	25	30	0.346	0.03	0.986	0.37	0.07	0.004	0.1	0.05	0.3	0.13	0.08	0.08	0.52	0.96	0.34
Delhi	8	7	0	0.12	0.993	0.57	0	0.044	0.25	0.14	0.37	0	0.13	0.13	0.25	0.88	0.54
HongKong	17	18	0.229	0.04	0.99	0.4	0.04	0.006	0.13	0.07	0.37	0.07	0.06	0.06	0.47	0.94	0.23
Lisbon	11	11	0.181	0.09	0.993	0.52	0	0.04	0.2	0.11	0.34	0.06	0.09	0.09	0.36	0.91	0.46
London	83	121	0.455	0.01	0.966	0.24	0.1	0.001	0.04	0.02	0.27	0.24	0.06	0.07	0.69	0.99	0.44
Lyon	10	10	0.192	0.11	0.994	0.53	0	0.048	0.22	0.12	0.36	0.07	0.1	0.1	0.4	0.9	0.53
Madrid	48	79	0.5	0.03	0.988	0.32	0.13	0.003	0.07	0.04	0.25	0.35	0.08	0.1	0.67	0.98	0.77
Marseille	6	5	0	0.17	0.995	0.62	0	0.073	0.33	0.18	0.45	0	0.17	0.17	0.33	0.83	0.79
MexicoCity	35	52	0.465	0.03	0.989	0.36	0.1	0.007	0.09	0.05	0.27	0.28	0.09	0.06	0.6	0.97	0.6
Milan	14	15	0.251	0.06	0.99	0.45	0.07	0.013	0.16	0.1	0.33	0.09	0.07	0.07	0.43	0.93	0.41
Montreal	10	10	0.192	0.11	0.994	0.53	0	0.048	0.22	0.12	0.36	0.07	0.1	0.1	0.4	0.9	0.53
Moscow	41	62	0.471	0.03	0.983	0.35	0.09	0.005	0.08	0.04	0.25	0.29	0.07	0.07	0.61	0.98	0.59
NewYork	77	109	0.443	0.01	0.971	0.25	0.05	0.001	0.04	0.02	0.28	0.22	0.06	0.04	0.68	0.99	0.29
Osaka	36	51	0.443	0.03	0.988	0.34	0.08	0.004	0.08	0.04	0.28	0.24	0.08	0.06	0.61	0.97	0.47
Paris	78	125	0.488	0.01	0.975	0.27	0.13	0.001	0.04	0.02	0.24	0.32	0.08	0.06	0.71	0.99	0.66
Prague	9	9	0.206	0.12	0.994	0.57	0.06	0.061	0.25	0.15	0.33	0.08	0.11	0.11	0.33	0.89	0.76
Rome	5	4	0	0.25	0.996	0.7	0	0.2	0.4	0.22	0.4	0	0.2	0.2	0.2	0.8	1.51
Seoul	71	111	0.48	0.01	0.98	0.26	0.09	0.001	0.04	0.02	0.27	0.3	0.08	0.07	0.76	0.99	0.46
Shanghai	22	28	0.389	0.04	0.989	0.41	0.05	0.01	0.12	0.07	0.28	0.18	0.09	0.09	0.55	0.95	0.45
Singapore	12	13	0.273	0.08	0.993	0.49	0.06	0.02	0.2	0.11	0.35	0.11	0.08	0.08	0.5	0.92	0.5
StPetersburg	14	16	0.317	0.07	0.992	0.49	0.07	0.026	0.18	0.1	0.31	0.13	0.07	0.07	0.43	0.93	0.56
Stockholm	20	19	0	0.02	0.981	0.34	0	0.003	0.1	0.05	0.4	0	0.05	0.05	0.4	0.95	0.05
Tokyo	62	107	0.512	0.02	0.985	0.31	0.15	0.002	0.06	0.03	0.23	0.39	0.08	0.06	0.71	0.98	0.88
Toronto	10	9	0	0.07	0.991	0.47	0	0.018	0.2	0.1	0.45	0	0.1	0.1	0.5	0.9	0.26
WashingtonDC	17	18	0.229	0.04	0.988	0.41	0.04	0.01	0.13	0.07	0.35	0.07	0.06	0.06	0.47	0.94	0.24

tion, and one single measure cannot fully capture the overall robustness of a metro network.



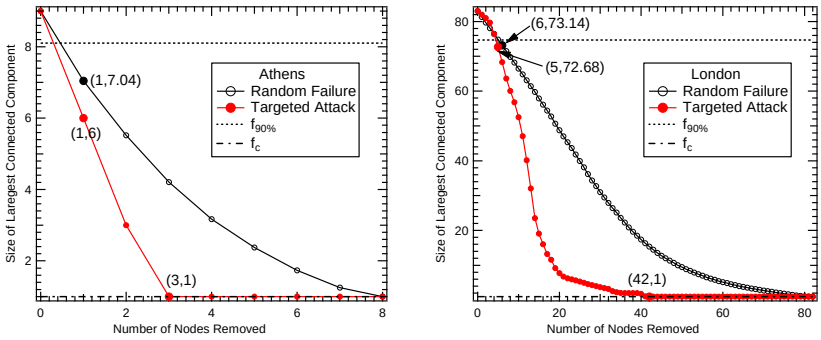
(a) Rome (b) Cairo and Marseille (c) Montreal (d) Toronto

Figure 4.3: The topology of metro networks.

Studying critical thresholds, Figure 4.4 shows the robustness level of metro net-

works, taking the *Athens* and *London* metro networks as examples, under random failures and deliberate attacks. The corresponding critical thresholds $f_{90\%}$ for targeted attacks (column 14) and random failures (column 15), and f_c for targeted attacks (column 16) and random failures (column 17) are shown in Table 4.1. Columns 14 and 15 in Table 4.1 show similar behavior of $f_{90\%}$ for targeted attacks and random failures.

Similar to the effective graph conductance C_G, *Rome* has the highest robustness level with $f_{90\%} = 0.20$ both for targeted attacks and random failures. *Cairo* and *Marseille* have the second highest robustness level with $f_{90\%} = 0.17$ for both targeted attacks and random failures. In contrast, and similar to the robustness indicator r^T, an evaluation of the critical threshold f_c under targeted attacks shows that *Seoul* and *Tokyo* are the most robust networks. *Seoul* has a critical threshold $f_c = 0.76$ indicating that 76% of nodes need to be removed before the network collapses. The critical threshold f_c under random failures shows that *London*, *NewYork*, *Paris* and *Seoul* are the most robust networks.



(a) Athens (b) London

Figure 4.4: Critical thresholds in metro networks under nodes removal.

4.4.2. METRIC CORRELATIONS

To assess the performance of theoretical metrics in capturing robustness, the Pearson correlation ρ between the ten robustness metrics and the critical thresholds in the metro networks is investigated. Moreover, the correlations within the ten robustness metrics are studied.

CORRELATION BETWEEN THEORETICAL AND NUMERICAL ROBUSTNESS METRICS

Table 4.2 presents the Pearson correlation between ten theoretical metrics and critical thresholds. The correlations between C_G and $f_{90\%}$ for random failures and targeted attacks are 0.89 and 0.91, respectively. The high positive correlation indicates that C_G effectively captures the 10% failure of the metro networks under node removal. Moreover, $E[\frac{1}{H}]$ and $\overline{\mu_{N-1}}$ also characterize the 10% failure of metro networks with performance slightly lower than C_G. The reliability Rel_G positively, but less strongly, correlates with critical thresholds $f_{90\%}$. However, the above mentioned metrics negatively correlate with f_c ($\rho(C_G, f_c) = -0.82$ for targeted attacks and $\rho(C_G, f_c) = -0.97$ under random failures).

Table 4.2: Pearson correlation ρ between theoretical robustness metrics and the critical thresholds.

	$f_{90\%}$-Degree	$f_{90\%}$-Random	f_c-Degree	f_c-Random
$\overline{r^T}$	-0.41	-0.52	0.87	0.85
C_G	0.89	0.91	-0.82	-0.97
Rel_G	0.54	0.59	-0.72	-0.75
$E[\frac{1}{H}]$	0.76	0.81	-0.9	-0.96
CC_G	-0.41	-0.52	0.73	0.66
$\overline{\mu_{N-1}}$	0.86	0.85	-0.71	-0.85
$\overline{E[D]}$	0.83	0.87	-0.87	-0.99
$\overline{\lambda}^*$	0.81	0.85	-0.88	-0.98
$1/\kappa$	0.56	0.64	-0.74	-0.83
M_G	-0.43	-0.53	0.89	0.8

The high correlation between $\overline{r^T}$ and f_c shows that $\overline{r^T}$ effectively characterizes when the network collapses under node removal. One explanation for the high correlation between $\overline{r^T}$ and f_c is that the robustness indicator $\overline{r^T}$ and the critical threshold f_c both characterize the number of alternative paths. Besides $\overline{r^T}$, the correlations of metrics M_G and CC_G to f_c suggest that these metrics have comparable performance in capturing when the network collapses. Yet, the correlations of $\overline{r^T}$, M_G and CC_G to $f_{90\%}$ are negative.

Metrics that positively correlate with $f_{90\%}$ and those that positively correlate with f_c therefore capture different aspects of metro networks as hinted above, and

both are important for robustness. Redundancy and contradiction between theoretical metrics are observed when capturing robustness of metros under node removal. Redundancy means that more than one metric positively correlates with critical thresholds and contradiction means that one specific metric positively correlates to $f_{90\%}$ while negatively correlates to f_c and vice versa.

CORRELATION WITHIN THEORETICAL ROBUSTNESS METRICS

To analyse the redundancy and contradiction of metrics, the Pearson correlation ρ between all the theoretical robustness metrics is investigated in Figure 4.5. In Figure 4.5, C_G, $E[\frac{1}{H}]$, $\overline{\mu_{N-1}}$ and Rel_G that effectively capture the critical threshold $f_{90\%}$ show a higher mutual correlation (e.g. $\rho\left(C_G, E[\frac{1}{H}]\right) = 0.95$). Similarly, for metrics $\overline{r^T}$, M_G and CC_G that capture the critical threshold f_c, a higher mutual correlation result is observed (e.g. $\rho\left(\overline{r^T}, CC_G\right) = 0.84$).

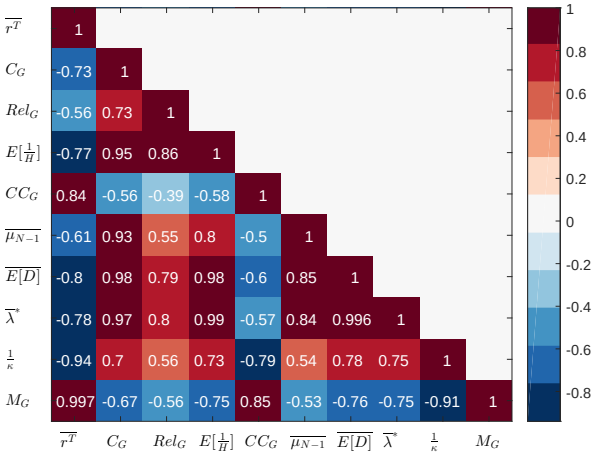


Figure 4.5: Pearson correlation ρ between theoretical robustness metrics.

As shown in Figure 4.5, these robustness metrics have a higher mutual correlation which indicates redundancy in capturing the robustness. Correspondingly, a representative set of robustness metrics by including only one metric from the mutually strongly depend set of metrics tends to sufficiently and effectively characterize the robustness [115]. For example, when quantifying the robustness $f_{90\%}$, including C_G in the representative set is more sufficient and effective than including C_G, $\overline{\mu_{N-1}}$ and $E[\frac{1}{H}]$.

In contrast to the positive and high correlations between certain metrics, the negative correlations in Figure 4.5 (e.g. -0.73 between $\overline{r^T}$ and C_G) might be problematic. In particular, when a higher $f_{90\%}$ and f_c are desired in the design of a metro, optimizing, for instance, both the robustness metrics $\overline{r^T}$ and C_G is beyond reach. Because maximizing $\overline{r^T}$ minimizes C_G and vice versa. This is therefore a major issue, which is not atypical of any robustness study. Indeed, while it is easy to develop design recommendations that can make a system more robust to certain conditions, it is much more challenging to develop recommendations that can make a system more robust overall. This point emphasizes the need to use multiple criteria when assessing the design of metro networks. It also points to the fact that robustness (and *resilience* more generally) are terms that are difficult to define and that cannot be solved with a simple objective function within an operation research context [127]. Instead, much work remains to be done to successfully come up with clear guidelines to transit planners, and simulation and network science may play an important part towards that end.

A possible approach to deal with this issue is suggested by Van Mieghem *et al.* [128], who defined a R-value, which is a weighted sum of all the considered theoretical metrics, i.e., R-value $= \sum_{i=1}^{M} w_i m_i$, where w_i is the weight for each metric m_i and M is the number of metrics taken into account. In the next subsection, we discuss another approach, which is based upon radar diagrams that are commonly used in urban planning and geography.

4.4.3. OVERALL ROBUSTNESS

To combine the ten calculated theoretical metrics that capture different aspects of robustness, we choose to draw radar diagrams for each metro. A radar diagram (also called star or spider diagram) is plot with as many axes as there are metrics, and the overall performance is calculated by measuring the area of the polygon formed. This type of diagram is especially useful when it is not possible to assign weights to individual metrics. First, for each set of metrics, each individual value x_i is being rescaled to a value in the interval $[0,1]$ using the rescaling formula: $(x_i - x_{min})/(x_{max} - x_{min})$. In the radar diagram, the robustness metrics are placed in a clockwise order. Metrics that are positively correlated with the critical threshold $f_{90\%}$ are located on one side and metrics that are negatively correlated to the critical

threshold $f_{90\%}$ are placed on the other side.

Figure 4.6 shows the radar plots for the 33 metro networks[6]. Moreover, Table 4.1 (last column) contains the areas of the polygons calculated. Overall, we can see that *Rome* and *Tokyo* are the top two of the most robust networks. Tokyo has many transfer stations in the periphery of the network that both enables it to offer many alternative paths and keep a relatively low resistance, hence ensuring a robust system. At the other hand of the spectrum, *Stockholm*, *Boston* and *Hong Kong* (the three least robust metros) have extensive networks with few transfer stations that inherently affect their robustness. Even *Washington DC* does not perform well because the transfer stations tend to be located in the city center, and it therefore achieves poorly in terms of "resistance" (i.e., long many stations without transfer from the terminals in the suburbs to the city center).

Most other networks tend to perform somewhat in between. From Figure 4.6, networks with polygons that are large in the bottom right corner tend to have many alternative paths. In contrast, metros with polygons that are large in the left-hand side tend to perform well in terms of resistance (as is the case for *Rome* despite its simple topology). *Mexico City* and *Berlin* deserve special attention since they seem to perform well in nearly all dimensions. *Berlin* has a particularly dense U-Bahn system, and *Mexico City* is known to have L-shaped lines to favor transferring [129].

From this work, clear recommendations can be set to promote a robust metro:

- Transfer stations are desirable to offer alternative paths. However, although large hubs are desirable to facilitate transferring, smaller hubs are as desirable to offer more options to transfer, thus offering more alternative paths (moreover they are less vulnerable to targeted attacks than large hubs).

- Long line sections are undesirable since a failure on one station will affect many passengers, likely resulting in the need for an emergency bus service to substitute failed stations. Transfer stations can therefore be located strategically to offer alternative paths while ensuring that line segments without transfer stations are kept as short as possible.

[6]The degree diversity κ instead of $\frac{1}{\kappa}$ is used in the radar diagram for the simplicity of computing the area.

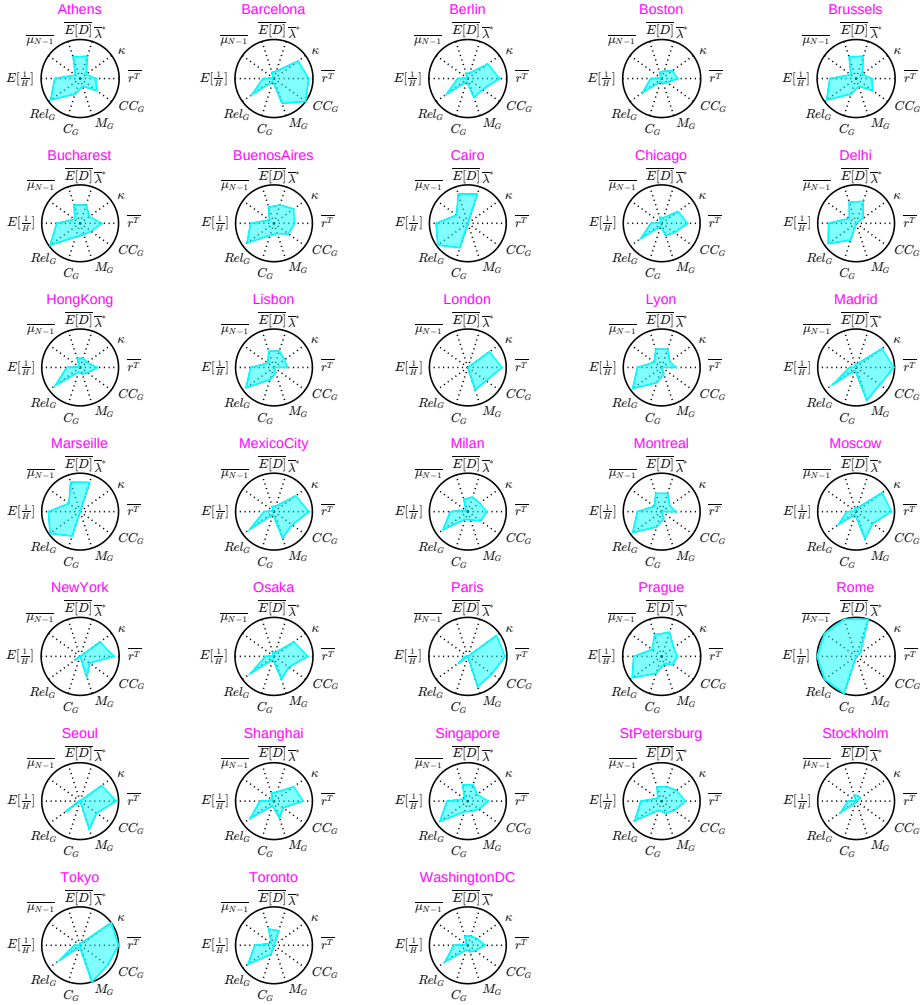


Figure 4.6: Radar diagrams for the 33 Metro Networks

4.5. CONCLUSION

The main objective of this work was to investigate the robustness of metro networks by analyzing several robustness metrics. In particular, we study ten theoretical robustness metrics and four numerical metrics. For the latter, we investigated two critical thresholds f, when 90% of the network is still remaining, $f_{90\%}$, and when the complete network is disintegrated, f_c (both under random failure and targeted attack).

Overall, we find that the ten theoretical robustness metrics capture two distinct aspects of the robustness of metro networks. A first aspect deals with the number of alternative paths, suggesting that more alternative paths is more desirable, as captured in r^T. In contrast, the second aspect deals with “resistance", suggesting that longer lines with no shorter alternative paths perform poorly, as captured in C_G. Essentially, as metro networks are expanded, effort should be put into creating transfer stations, both in city centers and peripheral areas to ensure that not only many alternative paths are created to reach a destination, but also that the average number of stations between two transfers is kept to a minimum. Overall we found that *Rome* benefits from shorter transferring paths and *Tokyo* are able to accomplish more transferring options.

Based on these observations and to fully capture these two aspects and assess the robustness of metro networks, we plotted the ten theoretical measures (standardized) on radar plots. This method offers both an equal representation of the variables at play as well as aesthetically-pleasing visual aid to help planners in their task to design robust metro networks.

Part II: Fundamentals of Graph Theory

5

DEGREE DISTRIBUTION AND ASSORTATIVITY IN LINE GRAPHS

5.1. INTRODUCTION

Infrastructures, such as the Internet, electric power grids and transportation networks, are crucial to modern societies. Most researches focus on the robustness of such networks to node failures [22, 130]. Specifically, the effect of node failures on the robustness of networks is studied by percolation theory both in single networks [22] and interdependent networks that interact with each other [35]. However, links frequently fail in various real-world networks, such as the failures of transmission lines in electrical power networks, path congestions in transportation networks. The concept of a line graph, that transforms links of the original graph into nodes in the line graph, can be used to understand the influence of link failures on infrastructure networks.

An undirected graph with N nodes and L links can be denoted as $G(N, L)$. The line graph $l(G)$ of a graph G is a graph in which every node in $l(G)$ corresponds to a link in G and two nodes in $l(G)$ are adjacent if and only if the corresponding links in G have a node in common [42]. The graph G is called the original graph of $l(G)$.

Line graphs are applied in various complex networks. Krawczyk *et al.* [131]

propose the line graph as a model of social networks that are constructed on groups such as families, communities and school classes. Line graphs can also represent protein interaction networks where each node represents an interaction between two proteins and each link represents pairs of interaction connected by a common protein [132]. By the line graph transformation, methodologies for nodes can be extended to solve problems related to links in a graph. For instance, the link chromatic number of a graph can be computed from the node chromatic number of its line graph [133]. Evan *et al.* [134] use algorithms that produce a node partition in the line graph to achieve a link partition in order to uncover overlapping communities of a network. Wierman *et al.* [135] improve the bond (link) percolation threshold of a graph by investigating site (node) percolation in its line graph.

Previous studies focus on various mathematical properties of line graphs. Whitney's Theorem [136] states that, if line graphs of two connected graphs G_1 and G_2 are isomorphic, the graphs G_1 and G_2 are isomorphic unless one is the complete graph K_3 and the other one is the star $K_{1,3}$. Krausz [137], Van Rooij and Wilf [138] have investigated the conditions for a graph to be a line graph. Van Rooij and Wilf [138] have studied the properties of graphs obtained by iterative usage of the line graph transformation, e.g., the line graph $l(G)$ of a graph G, the line graph $l(l(G))$ of the line graph $l(G)$, etc. Furthermore, Harary [139] has shown that for connected graphs that are not path graphs, all sufficiently high numbers of iterations of the line graph transformation produce Hamiltonian graphs[1]. The generation of a random line graph is studied in [140]. An original graph can be reconstructed [141–143] from its line graph with a computational complexity that is linear in the number of nodes N.

In this chapter, we analytically study the degree distribution and the assortativity of line graphs and the relation to the degree distribution and the assortativity of their original networks. We show that the degree distribution in the line graph of the Erdős-Rényi graph follows the same pattern as the degree distribution in Erdős-Rényi. However, the line graph of an Erdős-Rényi graph is not an Erdős-Rényi graph. Additionally, we investigate the assortativity of line graphs and show that the assortativity is not linearly related to the assortativity in the original graphs. The line

[1] A Hamiltonian graph is a graph possessing a Hamiltonian cycle which is a closed path through a graph that visits each node exactly once.

graphs are assortative in most cases, yet line graphs are not always assortative. We investigate graphs with negative assortativity in their line graphs. The remainder of this chapter is organized as follows. The degree distribution of line graphs is presented in Section 5.2. Section 5.3 provides the assortativity of line graphs. We conclude in Section 5.4.

5.2. DEGREE DISTRIBUTION

Random graphs are developed as models of real-world networks of several applications, such as peer-to-peer networks, the Internet and the World Wide Web. The degree distribution of Erdős-Rényi random graphs and scale free graphs are recognized by the binomial distribution and the power law distribution, respectively. This section studies the degree distribution of the line graphs of Erdős-Rényi and scale free graphs.

Let $G(N,L)$ be an undirected graph with N nodes and L links. The adjacency matrix A of a graph G is an $N \times N$ symmetric matrix with elements a_{ij} that are either 1 or 0 depending on whether there is a link between nodes i and j or not. The degree d_i of a node i is defined as $d_i = \sum_{k=1}^{N} a_{ik}$. The degree vector $d = (d_1\ d_2\ \cdots\ d_N)$ has a vector presentation as $Au = d$, where $u = (1,1,\cdots,1)$ is the all-one vector. The adjacency matrix [42] of the line graph $l(G)$ is $A_{l(G)} = R^T R - 2I$, where R is an $N \times L$ unsigned incidence matrix with $R_{il} = R_{jl} = 1$ if there is a link l between nodes i and j, elsewhere 0 and I is the identity matrix. The degree vector $d_{l(G)}$ of the line graph $l(G)$ is $d_{l(G)} = A_{l(G)} u_{L\times 1}$. For an arbitrary node l in the line graph $l(G)$, which corresponds to a link l connecting nodes i and j in graph G (as shown in Figure 5.1), the degree d_l of the node l follows

$$d_l = d_i + d_j - 2 \tag{5.1}$$

The random variable D_i denotes the degree of a randomly chosen node i in Erdős-Rényi graphs $G_p(N)$ and (5.1) shows that the degree D_l of a link l with end node i in the corresponding line graph is $D_l = D_i + D_j - 2$.

Theorem 4. *The degree distribution of the line graph $l(G_p(N))$ of an Erdős-Rényi*

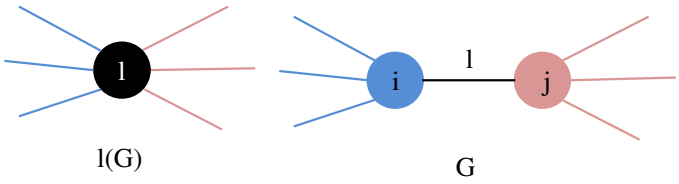


Figure 5.1: Node l in line Graph $l(G)$ corresponds to the link l in G.

graph $G_p(N)$ follows a binomial distribution

$$Pr[D_l = k] = \binom{2N-4}{k} p^k (1-p)^{(2N-4-k)} \tag{5.2}$$

with average degree $E[D_{l(G_p(N))}] = (2N-4)p$.

5

Proof. Applying (5.1), the degree distribution D_l of a node l in a line graph is

$$\Pr[D_l = k] = \Pr[D_i + D_j - 2 = k]$$

Using the law of total probability [33] yields

$$\Pr[D_l = k] = \sum_{m=1}^{k} \Pr[D_j = k - m + 2 \mid D_i = m] \Pr[D_i = m]$$

Since the random variables D_i and D_j in $G_p(N)$ are independent, we have

$$\Pr[D_l = k] = \sum_{m=1}^{k} \Pr[D_j = k - m + 2] \Pr[D_i = m] \tag{5.3}$$

An arbitrarily chosen (i.e. uniformly at random) node l in the line graph $l(G)$ corresponds to an arbitrarily chosen link in G. The degree distribution [33] of the end node i of an arbitrarily chosen link in G is

$$\Pr[D_i = m] = \frac{m \Pr[D = m]}{E[D]} \tag{5.4}$$

where $\Pr[D = m]$ is the degree distribution of an arbitrarily chosen node in graph G and $E[D]$ is the average degree of an arbitrarily chosen node. In an Erdős-Rényi graph, we have $\Pr[D = m] = \binom{N-1}{m} p^m (1-p)^{N-1-m}$ and $E[D] = (N-1)p$. By substituting (5.4) into (5.3) and applying the binomial distribution of random variables D_i

and D_j, we have

$$\begin{aligned}\Pr[D_l = k] &= \sum_{m=1}^{k} \frac{(k-m+2)\Pr[D = k-m+2]}{E[D]} \frac{m\Pr[D=m]}{E[D]} \\ &= \sum_{m=1}^{k} \frac{(k-m+2)\binom{N-1}{k-m-2} p^{k-m+2}(1-p)^{N-1-(k-m+2)}}{(N-1)p} \frac{m\binom{N-1}{m} p^m (1-p)^{N-1-m}}{(N-1)p} \\ &= p^k (1-p)^{2N-4-k} \sum_{m=0}^{k} \binom{N-2}{k-m}\binom{N-2}{m}\end{aligned}$$

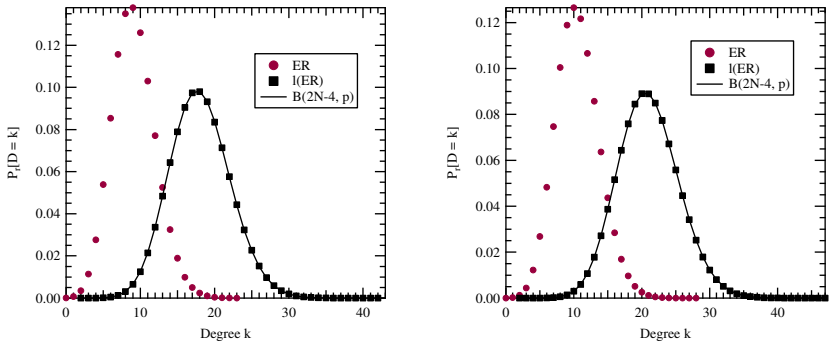
Using Vandermonde's identity $\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}$, we arrive at (5.2). □

Theorem 4 illustrates that the degree distribution of the line graph $l(G)$ of an Erdős-Rényi graph G follows a binomial distribution with average degree $E[D_{l(G)}] = (2N-4)p$. Compared to the average degree $E[D] = (N-1)p$, the average degree of the line graph of the Erdős-Rényi graph is two times the average degree $E[D]$ of the Erdős-Rényi graph minus $2p$.

Figure 5.2 shows the degree distribution of the line graphs of Erdős-Rényi graphs $G_N(p)$ for $N = 100$, 200 and $p = 2p_c$ ($p_c \approx \frac{\ln N}{N}$), where 10^5 Erdős-Rényi graphs are generated. In Figures 5.2(a) and (b), the degree distributions of Erdős-Rényi graphs (red circle) follow a binomial distribution. The degree distribution of the corresponding line graph (black square) is fitted by a binomial distribution $B(2N-4, p)$. The simulation results agree with Theorem 4. Moreover, the average degree $E[D_{l(G)}]$ of the line graph is approximately two times the average degree $E[D]$ of the graph G.

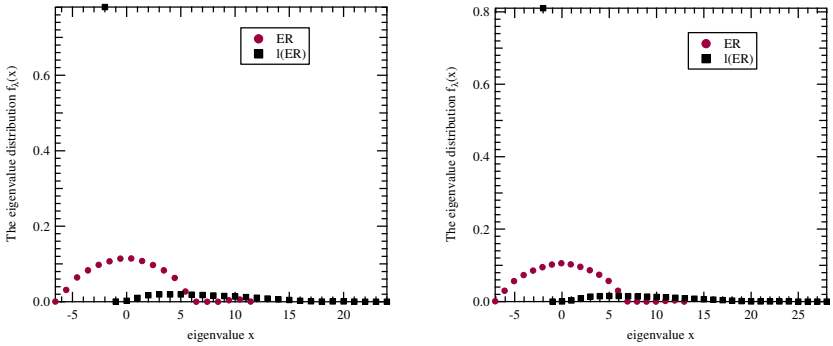
Since the degree distribution of the line graphs of Erdős-Rényi graphs follows a binomial distribution, we pose the question: Is the line graph of an Erdős-Rényi graph also an Erdős-Rényi graph? In order to answer this question, we investigate the eigenvalue distribution of the line graph. Figure 5.3 shows the eigenvalue distribution of Erdős-Rényi graphs and their line graphs. As shown in [42], the eigenvalue distribution of Erdős-Rényi graphs follows a semicircle distribution. The eigenvalue distribution of the line graphs of Erdős-Rényi graphs follows a different distribution than a semicircle distribution. Since the spectrum of a graph can be regarded as the unique fingerprint of that graph to a good approximation [144], we conclude that the line graphs of Erdős-Rényi graphs are not Erdős-Rényi graphs.

Generating functions are powerful to study the degree distribution of networks



(a) $N = 100$, $p = 2p_c$ (b) $N = 200$, $p = 2p_c$

Figure 5.2: The degree distribution of Erdős-Rényi graphs and their corresponding line graphs.



(a) $N = 100$, $p = 2p_c$ (b) $N = 200$, $p = 2p_c$

Figure 5.3: The eigenvalue distribution of Erdős-Rényi graphs and their corresponding line graphs. The simulations are performed on 10^5 instances.

[33]. Assuming the degree independence of nodes in graph G, Theorem 5 shows the generating function for the line graph $l(G)$ of an arbitrary graph.

Theorem 5. *Assuming that the degrees of nodes in a graph G are independent, the generating function for the degree D_l in the line graph $l(G)$ follows*

$$\varphi_{D_l}(z) = \left(\frac{E[z^{D_{l^+}}]}{z}\right)^2 \tag{5.5}$$

where D_{l^+} is the degree of the end node of an arbitrarily chosen link l in G.

Proof. The probability generating function for the degree D_l of a node l in the line

graph is

$$\varphi_{D_i}(z) = E[z^{D_l}]$$

Using (5.1), we have

$$\varphi_{D_i}(z) = E[z^{D_i+D_j-2}]$$

Since the condition in the theorem assumes that the random variables D_i and D_j are independent and identically distributed as D_{l^+}, we establish Theorem 5. □

We apply the generating function (5.5) in the line graph whose original graph has a power law degree distribution with the exponent γ, and has independent nodal degrees. In Appendix B, we deduce that, with $\gamma_G = 3$ in the original graph,

$$\Pr[D_l = k] \propto \left(\frac{1}{k+2}\right)^{\gamma_{l(G)}} \tag{5.6}$$

where $\gamma_{l(G)} = 2$. Equation (5.6) illustrates that, when we assume that the degrees in the original graph are independent, the degree distribution in the line graph follows a power law degree distribution. However, due to the preferential attachment in scale-free graphs and $2L = \sum_{i=1}^{N} d_i$, the node degrees are dependent rather than independent. Correspondingly, a gap is observed in Figure 5.4 between the approximation equation (5.6) (blue circle) and the simulation result (red square).

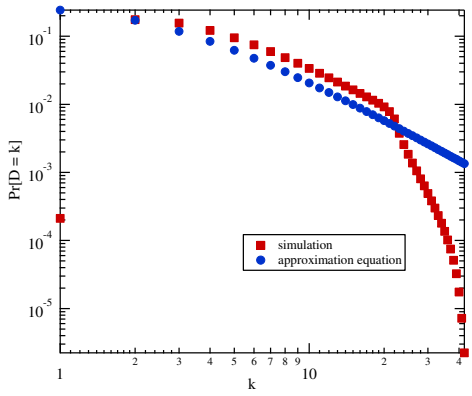


Figure 5.4: The degree distribution in the line graph of the Barabási-Albert graph both from simulations and the approximation equation (5.6). Both the x-axis and the y-axis are in log scale. The simulations are performed on 10^5 Barabási-Albert graphs with $N = 500$ and average degree 4. The cut-off in the simulation is due to the finite size of the Barabási-Albert graph.

The dependency assumption in (5.5) can be assessed by the total variation distance $d_{TV}(X,Y)$, defined as [33]:

$$d_{TV}(X,Y) = \sum_{k=-\infty}^{\infty} \big|\Pr[X=k] - \Pr[Y=k]\big|$$

where $\Pr[X=k]$ denotes the probability density function for (5.6) and $\Pr[Y=k]$ for simulations.

Figure 5.5 shows the total variation distance when the number of nodes N in Barabási-Albert graphs increases from 500 to 1000 with average degree 4. For each size of the original graph, 10^5 graphs are generated. Figure 5.5 demonstrates that $d_{TV}(X,Y)$ decreases with the number of nodes N, starting from 0.667 when $N=500$ to 0.640 when $N=1000$. Accordingly, the accuracy of the approximation equation (5.6) increases with the size of the original graph.

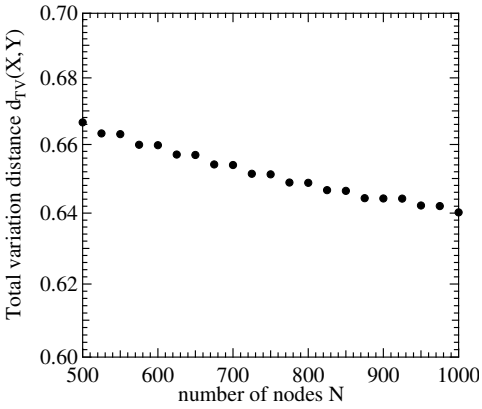


Figure 5.5: The total variation distance $d_{TV}(X,Y)$ when the original graph has different number of nodes from 500 to 1000.

5.3. ASSORTATIVITY

Networks with a same degree distribution may have significantly different topological properties [145]. Networks, where nodes preferentially connect to nodes with (dis)similar property, are called (dis)assortative [146]. An overview of the assortativity in complex networks is given in [147]. Assortativity is quantified by the linear degree correlation coefficient defined as

$$\rho_{D_{l(G)}} = \frac{E[D_{l^+}D_{l^-}] - E[D_{l^+}]E[D_{l^-}]}{\sigma_{D_{l^+}}\sigma_{D_{l^-}}} \tag{5.7}$$

where $E[X]$ and σ_X are the mean and standard deviation of the random variable X. The definition (5.7) has been transformed into a graph formulation in [145]. In this section, we investigate the assortativity $\rho_{D_{l(G)}}$ of the line graph $l(G)$ and its relation to the assortativity of the graph G.

5.3.1. ASSORTATIVITY IN THE LINE GRAPH

In this subsection, we derive a formula for the assortativity in a general line graph, represented in Theorem 6. The relation between the assortativity in the line graph and the assortativity in the original graph is shown in Corollary 1.

Theorem 6. *The assortativity in the line graph $l(G)$ of a general graph G is*

$$\rho_{D_{l(G)}} = 1 - \frac{d^T A\Delta d - N_4}{3d^T A\Delta d + \sum_{k=1}^{N} d_k^4 - 2\sum_{k=1}^{N} d_k^3 - 2N_3 - \frac{(N_3+\sum_{k=1}^{N} d_k^3 - 2N_2)^2}{N_2 - N_1}}$$

where d is the degree vector, $\Delta = diag(d_i)$ is the diagonal matrix with the nodal degrees in G and $N_k = u^T A^k u$ is the total number of walks of length k.

The proof for Theorem 6 is given in Appendix B.2. In order to investigate the relation between the assortativity of the line graph $l(G)$ and the assortativity of the graph G, Corollary 1 rephrases the assortativity $\rho_{D_{l(G)}}$ of the line graph $l(G)$ in terms of the assortativity ρ_D of the graph G.

Corollary 1. *The assortativity $\rho_{D_{l(G)}}$ of the line graph can be written in terms of the assortativity ρ_D of the graph G as*

$$\rho_{D_{l(G)}} = 1 - \frac{(d^T A\Delta d - N_4)\mu^2}{(N_2 - N_1)\left(-4(1+\rho_D)^2\left(\frac{1}{N_1}\sum_{i=1}^{N} d_i^3 - \left(\frac{N_2}{N_1}\right)^2\right)^2 + 2\mu^2(1+\rho_D)\left(\frac{1}{N_1}\sum_{i=1}^{N} d_i^3 - \left(\frac{N_2}{N_1}\right)^2\right) + \mu u_3\right)}$$

where $\mu = E[D_{l(G)}]$ and $u_3 = E[(D_{l(G)} - E[D_{l(G)}])^3]$.

The proof for Corollary 1 is given in Appendix B.3. Corollary 1 indicates that the assortativity of the line graph is not linearly related to the assortativity of the original graph. For the Erdős-Rényi graphs, a relatively precise relation between the assortativity of the line graph and the one of the original graph is given in Theorem 7.

Theorem 7. *The difference between the assortativity $\rho_{D_{l(G)}}$ of the line graph of an Erdős-Rényi graph $G_N(p)$ and the assortativity ρ_{D_G} of $G_N(p)$ converges to 0.5 in the limit of large graph size N.*

Proof. Based on the definition in equation (5.7) and denoting $l^+ = i \sim c$ and $l^- = c \sim j$, we have

$$\rho_{D_{l(G)}} = \frac{E[(D_i + D_c)(D_j + D_c)] - E[D_i + D_c]E[D_j + D_c]}{\sigma_{D_i+D_c}\sigma_{D_j+D_c}}$$
$$= \frac{E[D_iD_j] - E[D_i]E[D_j] + E[D_iD_c] - E[D_i]E[D_c] + E[D_jD_c] - E[D_j]E[D_c] + E[D_c^2] - E^2[D_c}{Var[D_i] + Var[D_c] + 2E[(D_i - E[D_i])(D_c - E[D_c])]}$$

In the connected Erdős-Rényi random graph in the limit of large graph size N, the assortativity ρ_{D_G} converges to zero [42] and we have

$$E[D_iD_j] - E[D_i]E[D_j] \approx 0$$

Similarly, $E[D_iD_c] - E[D_i]E[D_c] \approx 0$ and $E[D_jD_c] - E[D_j]E[D_c] \approx 0$. Combining with $E[(D_i - E[D_i])(D_c - E[D_c])] = E[D_iD_c] - E[D_i]E[D_c] \approx 0$, we arrive at

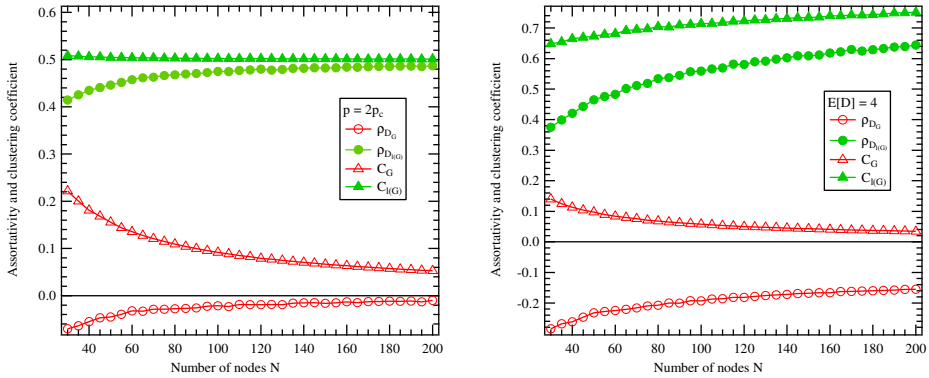
$$\rho_{D_{l(G)}} \approx \frac{E[D_c^2] - E^2[D_c]}{2\text{Var}[D_c]} = 0.5$$

□

In order to verify Theorem 7, Figure 5.6 shows the assortativity of (a) Erdős-Rényi graphs, (b) Barabási-Albert graphs, and the assortativity of their corresponding line graphs. In Figure 5.6(a), the assortativity of $G_p(N)$ converges to 0 with the increase of the graph size N. Correspondingly, the assortativity in the line graph of $G_p(N)$ converges to 0.5 which confirms Theorem 7. Based on the assortativity ρ_D of a connected Erdős-Rényi graph $G_p(N)$, which is zero [42, 146] in the limit of large graph size, we again verify that the line graph of an Erdős-Rényi graph is not an Erdős-Rényi graph. Figure 5.6(b) illustrates the assortativity $\rho_{D_{l(G)}}$ of the line graph of the Barabási-Albert graph is also positive and increases with the graph size.

Youssef *et al.* [148] show that the assortativity is related to the clustering coefficient[2] C_G. Specifically, assortative graphs tend to have a higher number $\blacktriangle_G$ of triangles and thus a higher clustering coefficient compared to disassortative graphs. Figure 5.6 shows that the assortative line graphs of both Erdős-Rényi and Barabási-Albert graph have a higher clustering coefficient (above 0.5). The results agree with the findings in [148].

[2]The clustering coefficient $C_G = \frac{3\blacktriangle_G}{N_2}$ is defined as three times the number $\blacktriangle_G$ of triangles divided by the number N_2 of connected triples.



(a) Erdős-Rényi graph.

(b) Barabási-Albert graph.

Figure 5.6: Assortativity ρ_D and clustering coefficient C_G of the (a) Erdős-Rényi graph $G_p(N)$ with $p = 2p_c$, (b) Barabási-Albert graph with the average degree $E[D] = 4$ and the corresponding line graph $l(G)$.

5

Table 5.1: Assortativity of real-world networks and their corresponding line graphs.

Networks	Nodes	Links	ρ_D	$\rho_{D_{l(G)}}$
Co-authorship Network [48]	379	914	−0.0819	0.6899
US airports [149]	500	2980	−0.2679	0.3438
Dutch Soccer [47]	685	10310	−0.0634	0.5170
Citation [150]	2678	10368	−0.0352	0.8127
Power Grid [17]	4941	6594	−0.0035	0.7007

Table 5.1 shows the assortativity of real-world networks and their corresponding line graphs. As shown in the table, the line graphs of all the listed networks show assortative mixing even though the original networks show dissortative mixing.

5.3.2. NEGATIVE ASSORTATIVITY IN LINE GRAPHS

Although the assortativity of a line graph is predominantly positive, we cannot conclude that the assortativity in any line graph is positive. This subsection presents graphs, whose corresponding line graphs possesses a negative assortativity.

THE LINE GRAPH OF A PATH GRAPH

A path graph P_N is a tree with two nodes of degree 1, and the other $N-2$ nodes of degree 2. The line graph $l(P)$ of a path graph P_N is still a path graph but with $N-1$ nodes. Observation 1 demonstrates that the assortativity in the line graph of a path

graph is always negative.

Observation 1. *The assortativity of the line graph $l(P)$ of a path P_N is*

$$\rho_{D_{l(P)}} = -\frac{1}{N-3}$$

where N is the number of nodes in the original path graph.

Proof. The reformulation [42] of the assortativity can be written as

$$\rho_D = 1 - \frac{\sum_{i\sim j}(d_i - d_j)^2}{\sum_{i=1}^{N-1}(d_i)^3 - \frac{1}{2L}(\sum_{i=1}^{N-1} d_i^2)^2} \tag{5.8}$$

Since the line graph of a path with N nodes is a path graph with $N-1$ nodes, where 2 nodes have node degree 1 and the other $(N-1)-2$ nodes have degree 2, we have that

$$\sum_{i=1}^{N-1} d_i^k = 2\times 1^k + ((N-1)-2)\times 2^k \tag{5.9}$$

and

$$\sum_{i\sim j}(d_i - d_j)^2 = 2\times 1^2 \tag{5.10}$$

Applying equations (5.9) and (5.10) into (5.8), we establish the Observation 1. □

The negative assortativity $\rho_{D_{l(P)}}$ of the line graph $l(P)$ of a path graph is an exception to the positive assortativity of the line graphs of the Erdős-Rényi graph, Barabási-Albert graph and real-world networks given in Table 5.1. Moreover, the assortativity of the line graph $l(P)$ is a fingerprint for the line graph $l(P)$ to be a path graph.

THE LINE GRAPH OF A PATH-LIKE GRAPH

Let $P^{m_1,\, m_2,\, \cdots,\, m_t}_{n_1,\, n_2,\, \cdots,\, n_t,\, p}$ be a path of p nodes $(1 \sim 2 \sim \cdots \sim p)$ with pendant paths of n_i links at nodes m_i, following the definition in [151]. We define the graph D_N through $D_N = P^2_{1,\, N-1}$ as drawn in Fig. 5.7. Observation 2 shows that the assortativity in the corresponding line graph $l(D_N)$ is always negative.

Observation 2. *The assortativity of the line graph $l(D_N)$ of the graph D_N in Figure 5.7 is*

$$\rho_{D_{l(D_N)}} = -\frac{1}{2N-3}$$

where N is the number of nodes in the graph D_N.

Figure 5.7: The graph D_N whose line graph has the negative assortativity.

Proof. Since 1 node has node degree 1, 1 node has node degree 3 and the other $(N-1)-2$ nodes have degree 2, we have that

$$\sum_{i=1}^{N-1} d_i^k = 1\times 1^k + 1\times 3^k + ((N-1)-2)\times 2^k \tag{5.11}$$

and

$$\sum_{i\sim j}(d_i - d_j)^2 = 1\times 1^2 + 3\times 1^2 \tag{5.12}$$

Applying equations (5.11) and (5.12) into (5.8), we establish the Observation 2. □

We define the graph E_N through $E_N = P^3_{1,\,N-1}$ as drawn in Fig. 5.8. The graph E_N is obtained from D_N by moving the pendant path from node 2 to node 3. The assortativity of the line graph $l(E_N)$ of the graph E_N is

$$\rho_{D_{l(E_N)}} = -\frac{1}{N-2}$$

For the graphs $P^{m_i}_{1,\,N-1}$ with one pendant path of 1 link at node m_i ($i = 2, 3, \cdots, N-2$),

Figure 5.8: The graph E_N whose line graph has the negative assortativity.

there are $N-3$ positions to attach the pendant path. Since the position for adding the pendant path is symmetric at $\lceil\frac{N-1}{2}\rceil$. We only consider i from 2 to $\lceil\frac{N-1}{2}\rceil$. Among all the graphs $P^{m_i}_{1,\,N-1}$ where $i = 2, 3, \cdots, \lceil\frac{N-1}{2}\rceil$), the line graphs of the graph D_N and E_N always have negative assortativity. The line graph of the graph $P^{m_i}_{1,\,N-1}$, where $i = 4, 5, \cdots, \lceil\frac{N-1}{2}\rceil$, has negative assortativity when the size N is small and has positive assortativity as N increases.

The graph $\widetilde{D}_N$ is defined through $\widetilde{D}_N = P^{2,\,N-3}_{1,\,1,\,N-2}$ as drawn in Fig. 5.9. Observation 3 shows that the assortativity in the corresponding line graph $l(\widetilde{D}_N)$ is always negative.

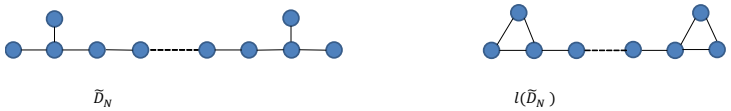


Figure 5.9: The graph $\widetilde{D}_N$ whose line graph has the negative assortativity.

Observation 3. *The assortativity of the line graph $l(\widetilde{D}_N)$ of the graph $\widetilde{D}_N$ in Figure 5.9 is*

$$\rho_{D_{l(\widetilde{D}_N)}} = -\frac{3}{N-3}$$

where N is the number of nodes in $\widetilde{D}_N$.

Proof. Since 2 nodes have node degree 3 and the other $(N-1)-2$ nodes have degree 2, we have that

$$\sum_{i=1}^{N-1} d_i^k = 2\times 3^k + ((N-1)-2)\times 2^k \tag{5.13}$$

and

$$\sum_{i\sim j}(d_i - d_j)^2 = 6\times 1^2 \tag{5.14}$$

Applying equations (5.13) and (5.14) into (5.8), we establish the Observation 3. □

The graphs $\widetilde{E}_N$ and $\widetilde{F}_N$ are defined through $\widetilde{E}_N = P^{2,\,N-4}_{1,\,1,\,N-2}$ and $\widetilde{F}_N = P^{3,\,N-4}_{1,\,1,\,N-2}$ as drawn in Fig. 5.10. The assortativity for the line graph of $\widetilde{E}_N$ is

$$\rho_{D_{l(\widetilde{E}_N)}} = -\frac{16}{5N-16}$$

The assortativity for the line graph of $\widetilde{F}_N$ is

$$\rho_{D_{l(\widetilde{F}_N)}} = -\frac{25}{7N-25}$$

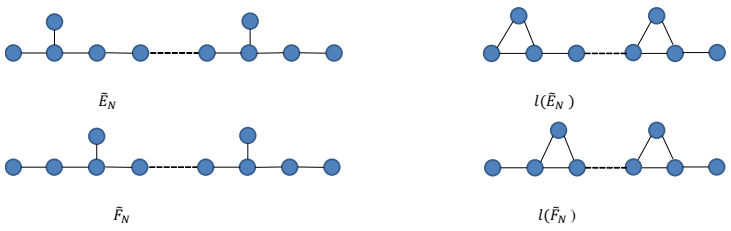


Figure 5.10: The graphs $\widetilde{E}_N$ and $\widetilde{F}_N$ whose line graphs have the negative assortativity.

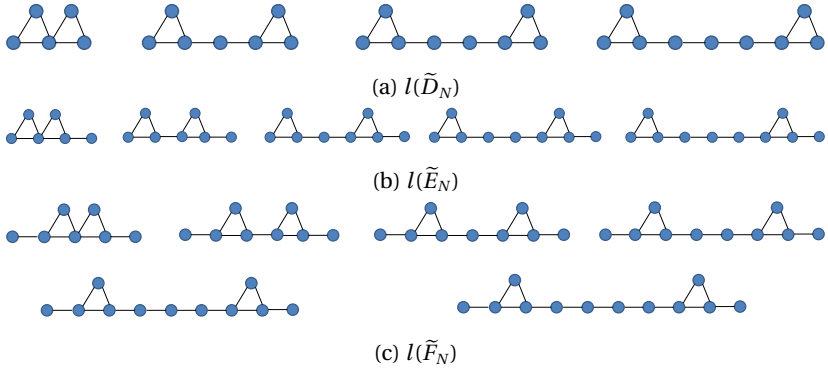
Graphs $\widetilde{D}_N$, $\widetilde{E}_N$, $\widetilde{F}_N$ are the graphs whose line graphs always have the negative assortativity. For the remaining graphs $P^{m_i,\, m_j}_{1,\, 1,\, N-2}$, $i \neq j$, their line graphs have negative assortativity when N is small. As N increases, the assortativity of the line graphs is positive.

LINE GRAPH OF NON-TREES

Both the path graphs and path-like graphs are trees. In this subsection, we study whether there exist non-trees whose line graphs have negative assortativity.

We start by studying the non-trees $l(D_N)$, $l(E_N)$ and $l(\widetilde{D}_N)$, $l(\widetilde{E}_N)$, $l(\widetilde{F}_N)$ in Figures 5.7-5.10. The non-tree graphs consist of cycles of 3 nodes connected by disjoint paths. The line graph of the non-tree $l(D_N)$ is denoted as $l(l(D_N))$, which is also the line graph of the line graph of D_N. By simulations we determine the non-tree graphs whose line graphs have negative assortativity. The results are given in Figures 5.11 and 5.12.

Figure 5.11: Non-tree graphs $l(D_N)$, $l(E_N)$ whose line graphs $l(l(D_N))$, $l(l(E_N))$ have negative assortativity.



(a) $l(\widetilde{D}_N)$

(b) $l(\widetilde{E}_N)$

(c) $l(\widetilde{F}_N)$

Figure 5.12: Non-tree graphs $l(\widetilde{D}_N)$, $l(\widetilde{E}_N)$, $l(\widetilde{F}_N)$ whose line graphs $l(l(\widetilde{D}_N))$, $l(l(\widetilde{E}_N))$, $l(l(\widetilde{F}_N))$ have negative assortativity.

As shown in Figures 5.11 and 5.12, for the line graphs of the non-trees to have negative assortativity, there can be either 1 or 2 cycles in the non-trees. In Figure 5.11, the line graph $l(l(E_N))$ of $l(E_N)$ has 1 cycle connected by two paths and the maximal path length is 2. In Figure 5.12, two cycles are connected by maximal 3

paths and the maximal path length is 4 in the line graph $l(l(\widetilde{F}_N))$. Moreover, for a line graph to have negative assortativity, the size of the original graph is in general small, less than 14 nodes in our simulations.

5.4. CHAPTER CONCLUSION

Topological characteristics of links influence the dynamical processes executed on complex networks triggered by links. The line graph, which transforms links from a graph to nodes in its line graph, generalizes the topological properties from nodes to links. This chapter investigates the degree distribution and the assortativity of line graphs. The degree distribution of the line graph of an Erdős-Rényi random graph follows the same pattern of the degree distribution as the original graph. We derive a formula for the assortativity of the line graph. We indicate that the assortativity of the line graph is not linearly related to the assortativity of the original graph. Moreover, the assortativity is positive for the line graphs of Erdős-Rényi graphs, Barabási-Albert graphs and most real-world networks. In contrast, certain types of trees, path and path-like graphs, have negative assortativity in their line graphs. Furthermore, non-trees consisting of cycles and paths can also have negative assortativity in their line graphs.

6

ORTHOGONAL EIGENVECTOR MATRIX OF THE LAPLACIAN

6.1. INTRODUCTION

Networks abound more than ever before. While many graph metrics have been proposed, that are reviewed e.g. in [152–154], the eigenvector structure of graph related matrices is hardly understood. A graph on N nodes can be represented by an $N \times N$ adjacency matrix A with $a_{ij} = 1$ if the pair of nodes is connected, otherwise $a_{ij} = 0$. Another graph related matrix is the Laplacian matrix $Q = \Delta - A$, where $\Delta = \text{diag}(d_i)$ is the $N \times N$ diagonal degree matrix and the degree of node i is $d_i = \sum_{j=1}^{N} a_{ij}$. When confining to an unweighted and undirected graph, the Laplacian matrix Q is symmetric and possesses the eigenvalue decomposition $Q = ZMZ^T$. The equality implies that all information at the left-hand side, that we call the topology domain, is also contained in the right-hand side, that we call the spectral domain. Most insight so far in graphs is gained in the topology domain that allows a straightforward drawing of a graph: nodes are interconnected by links and display a typical graph representation, attractive and understandable to humans. The spectral domain, consisting of the set $\{z_1, z_2, \ldots, z_N\}$of eigenvectors of the Laplacian Q and the corresponding set of eigenvalues in M, is less intuitive for humans. However, as mentioned in the

preface of [42], the spectral decomposition $Q = ZMZ^T$ (or $A = X\Lambda X^T$) represents a transformation of a similar nature as a Fourier transform, which suggests that some information is better or more adequately accessible in one domain and other information in the other domain.

Most spectral results are obtained for eigenvalues, and in particular the largest eigenvalue or spectral radius [155] for the adjacency matrix and the second smallest eigenvalue or the algebraic connectivity [43] for the Laplacian matrix. The spectral radius plays an important role in characterizing the dynamical process on networks, such as SIS (susceptible-infected-susceptible) epidemic spread [27]. The algebraic connectivity [43] plays an important role in bounding the node and link connectivity, i.e. the number of nodes and links that have to be removed to disconnect the graph. Correspondingly, the algebraic connectivity is considered as a robustness measure against node/link failures [156]. The sum of the inverse Laplacian eigenvalues, called the effective graph resistance [72], can be used to improve the robustness of complex networks [77].

While the number of mathematical results on other eigenvalues is already considerably less, results on eigenvectors are relatively scarce [157, 158]. Most results on eigenvectors focus on the principle eigenvector [53], the eigenvector corresponding to the largest eigenvalue of the adjacency matrix of a graph, or the Fiedler vector [43, 159], the eigenvector belonging to the second smallest eigenvalue of the Laplacian matrix.

Here, we approach the challenge of unravelling the "hidden information" in the orthogonal eigenvector matrix Z of the Laplacian matrix by extensive simulations, because the purely mathematical discovery of nice properties of the matrix Z seems of a daunting difficulty. Since many properties of the Erdős-Rényi (ER) graphs $G_p(N)$ are known [160], we concentrate here only on this class of graphs. An ER graph $G_p(N)$ on N nodes and with link density p is generated by randomly connecting a pair of nodes with a probability p, independently of any other pair. Although ER graphs are generally not good representatives of real-world networks, we believe that, if we cannot understand this simple class of random graphs, the more realistic (but more complex) classes of graphs are certainly beyond reach. Thus, here, we make a first step to learn about the properties of orthogonal eigenvector matrix Z of the Laplacian by confining to ER graphs. An extra bonus, apart from a

computational advantage, is that relatively small sizes N in the class $G_p(N)$, even below $N = 100$, already give a good reflection of the general properties for any N.

The chapter is organized as follows. Section 6.2 presents the definition and the orthogonality properties of the eigenvector matrix of the Laplacian. Section 6.3 illustrates the properties of the eigenvector matrix. The dual fundamental weight vector is introduced and the distribution of the dual fundamental weight is studied in Section 6.4. Section 6.5 concludes the chapter.

6.2. EIGENSTRUCTURE OF THE LAPLACIAN Q OF A GRAPH

As in [42], we denote by z_k the eigenvector of the $N \times N$ symmetric matrix Q belonging to the eigenvalue μ_k, normalized so that $z_k^T z_k = 1$. The eigenvalues of $Q = Q^T$ are real and can be ordered as $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_N$. The all-one vector $u = (1, 1, \ldots, 1)$ is the eigenvector belonging to $\mu_N = 0$, since the row sum is $Qu = 0$ for any Laplacian matrix. Let Z be the orthogonal matrix with the eigenvectors of Q in the columns,

$$Z = \begin{bmatrix} z_1 & z_2 & z_3 & \cdots & z_N \end{bmatrix}$$

or explicitly in terms of the m-th component $\left(z_j\right)_m$ of eigenvector z_j,

$$Z = \begin{bmatrix} (z_1)_1 & (z_2)_1 & (z_3)_1 & \cdots & (z_N)_1 \\ (z_1)_2 & (z_2)_2 & (z_3)_2 & \cdots & (z_N)_2 \\ (z_1)_3 & (z_2)_3 & (z_3)_3 & \cdots & (z_N)_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (z_1)_N & (z_2)_N & (z_3)_N & \cdots & (z_N)_N \end{bmatrix} \tag{6.1}$$

where the element $Z_{ij} = \left(z_j\right)_i$. The eigenvalue equation $Qz_k = \mu_k z_k$ translates to the matrix equation $Q = ZMZ^T$, where $M = \mathrm{diag}\left(\mu_k\right)$.

The relation $Z^T Z = I = ZZ^T$ (see e.g. [42, p. 223]) expresses, in fact, double orthogonality. The first equality $Z^T Z = I$ translates to the well-known orthogonality relation

$$z_k^T z_m = \sum_{j=1}^{N} (z_k)_j (z_m)_j = \delta_{km} \tag{6.2}$$

stating that the eigenvector z_k belonging to eigenvalue μ_k is orthogonal to any other eigenvector belonging to a different eigenvalue. The second equality $ZZ^T = I$, which

arises from the commutativity of the inverse matrix $Z^{-1} = Z^T$ with the matrix Z itself, can be written as $\sum_{j=1}^{N} \left(z_j\right)_m \left(z_j\right)_k = \delta_{mk}$ and suggests us to define the row vector in Z as

$$y_m = ((z_1)_m, (z_2)_m, \ldots, (z_N)_m) \tag{6.3}$$

Then, the second orthogonality condition $ZZ^T = I$ implies orthogonality of the vectors

$$y_l^T y_j = \sum_{k=1}^{N} (z_k)_l (z_k)_j = \delta_{lj} \tag{6.4}$$

The fundamental weight $\omega_k = u^T z_k$ and the dual fundamental weight $\varphi_j = u^T y_j$ have been introduced in [161]. The corresponding vectors $\omega = (\omega_1, \omega_2, \cdots, \omega_N)$ and $\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_N)$ can be written as the column sum and the row sum, respectively, of the orthogonal matrix Z

$$\omega = Z^T u \tag{6.5}$$

and

$$\varphi = Zu \tag{6.6}$$

6

Instead of concentrating on the adjacency matrix A, we consider here the Laplacian matrix Q, mainly because the all-one vector u is always an eigenvector of Q, which greatly simplifies the fundamental weight vector ω. Indeed, since the normalized Laplacian eigenvector $z_N = \frac{u}{\sqrt{N}}$ belonging to the smallest eigenvalue $\mu_N = 0$ is orthogonal to all other eigenvectors, it follows from (6.5) that, in a connected graph,

$$\omega = (0, 0, \cdots, \sqrt{N}) = \sqrt{N} e_N \tag{6.7}$$

6.3. EXPLORING PROPERTIES OF THE ORTHOGONAL EIGENVECTOR MATRIX Z OF THE LAPLACIAN Q

Via extensive simulations on Erdős-Rényi (ER) graphs $G_p(N)$, initial insight is gained in the sum of all the elements, the number of zero elements and the maximum and the minimum element in the eigenvector matrix Z of the Laplacian matrix Q.

6.3.1. THE SUM s_Z OF THE ELEMENTS IN Z

Let s_Z be the sum of the elements in the matrix Z. Using the definitions (6.5) and (6.6) for a connected graph, the sum $s_Z = u^T Z u = u^T \varphi$ as well as $s_Z = \left(Z^T u\right)^T u =$

$\omega^T u = \sqrt{N}$, where (6.7) has been used. In a disconnected graph G, the sum s_Z is

$$s_Z = \sum_{j=1}^{c} \sum_{k=1}^{N} (z_j)_k$$

where c is the number of components in the disconnected graph G. For the case $c = 2$, more details are discussed in the Appendix C.

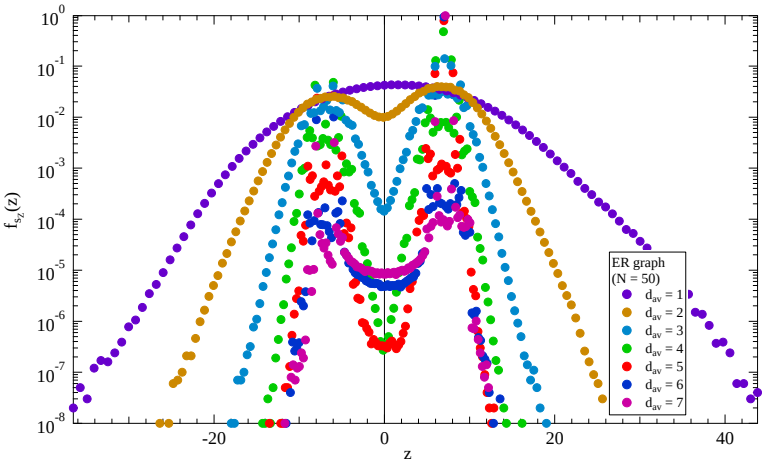


Figure 6.1: The probability density function of s_Z in ER random graphs $G_p(N)$ for $N = 50$ and various average degree $d_{av} = p(N-1)$, ranging from $d_{av} = 1$ up to $d_{av} = 7$. The y-axis is in log-scale.

Fig. 6.1 shows the probability density function $f_{s_Z}(z)$ in ER graphs $G_p(N)$ for $N = 50$ and various average degree $d_{av} = p(N-1)$, ranging from $d_{av} = 1$ up to $d_{av} = 7$. We have generated 10^8 ER graphs $G_p(50)$. Fig. 6.1 demonstrates that the maximum value of $f_{s_Z}(z)$ at $z = \sqrt{N}$ increases with the average degree d_{av}. For $d_{av} \geq 4$, the maximum value of $f_{s_Z}(z)$ is dominantly high because most generated graphs are connected. Indeed [33], for $N = 50$ and $d_{av} \approx 3.9$, $\Pr[G_p(N)$ is connected$]$ is about 36%. Moreover, ignoring the peak value at $z = \sqrt{N}$, we observe that $f_{s_Z}(z)$ is roughly symmetric around 0.

6.3.2. THE NUMBER z_Z OF ZERO ELEMENTS IN Z

The number of zero elements in the orthogonal matrix Z is an integer smaller than $N^2 - N$, because each eigenvector is different from the zero vector and, thus, should contain at least one non-zero element. Hence, $0 \leq z_Z \leq N^2 - N$. In the simulations, an element in Z with absolute value smaller than 10^{-10} is considered as

zero.

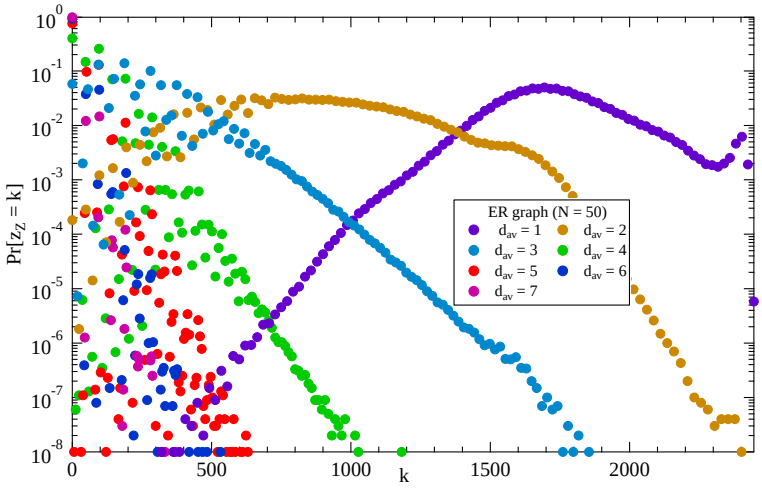


Figure 6.2: The probability $\Pr[z_Z = k]$ that the number of zeros in Z equals k in ER random graphs $G_p(N)$ for $N = 50$ and various average degree $d_{av} = p(N-1)$, ranging from $d_{av} = 1$ up to $d_{av} = 7$. The y-axis is in log-scale.

Fig. 6.2 shows that, in ER graphs of $N = 50$ nodes, the average number $E[z_Z]$ of zero elements decreases with the average degree d_{av}. The probability $\Pr[z_Z = 0]$ that there is no zero element increases with d_{av}. More specifically, for small average degrees, $d_{av} = 1$ and $d_{av} = 2$, the average number $E[z_Z]$ of zero elements is high and the probability that $\Pr[z_Z = 0]$ is small (and almost zero for $d_{av} = 1$). For $d_{av} \geq 4$, the probability $\Pr[z_Z = 0]$ is dominantly high. Moreover, only for $d_{av} \leq 3$, the curve $\Pr[z_Z = k]$ versus k is reasonably stable, but for $d_{av} \geq 4$, large scattering is observed.

6.3.3. THE MINIMUM AND MAXIMUM ELEMENT IN Z

We denote the minimum element in the orthogonal matrix Z by $\zeta_Z = \min_{ij} z_{ij}$ and the maximum element by $\xi_Z = \max_{ij} z_{ij}$.

Figs. 6.3 and 6.4 demonstrate that $\xi_Z \stackrel{d}{=} -\zeta_Z$, where $\stackrel{d}{=}$ denotes equality in distribution, which is less strong than $\max_{ij} z_{ij} = -\min_{ij} z_{ij}$. Fig. 6.4 indicates that the lower the average degree d_{av}, the higher the probability that the maximum ξ_Z attains the value 1. If only one element is non-zero, then that element must equal ± 1 because of the normalization of eigenvectors.

If the graph is connected, then $z_N = \frac{u}{\sqrt{N}}$ (else, there are c components leading to a different normalization of the u vector, see the Appendix C). The second orthog-

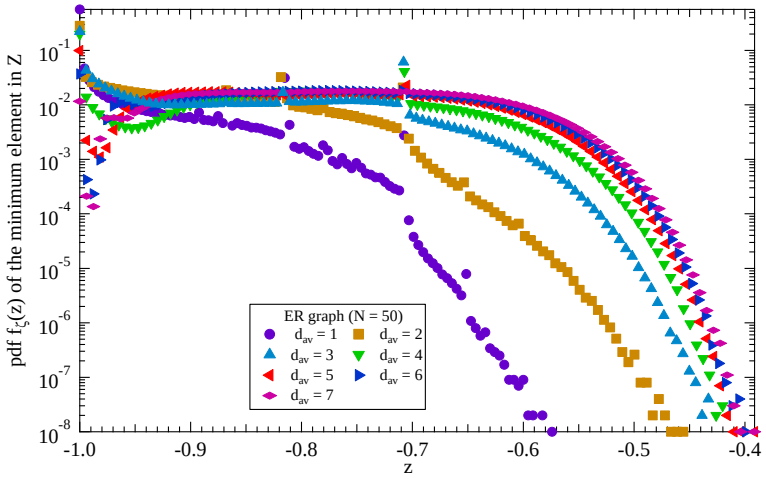


Figure 6.3: The probability density function $f_\zeta(z)$ of the minimum element in Z in ER random graphs $G_p(N)$ for $N = 50$ and various average degree $d_{av} = p(N-1)$, ranging from $d_{av} = 1$ up to $d_{av} = 7$. The y-axis is in log-scale.

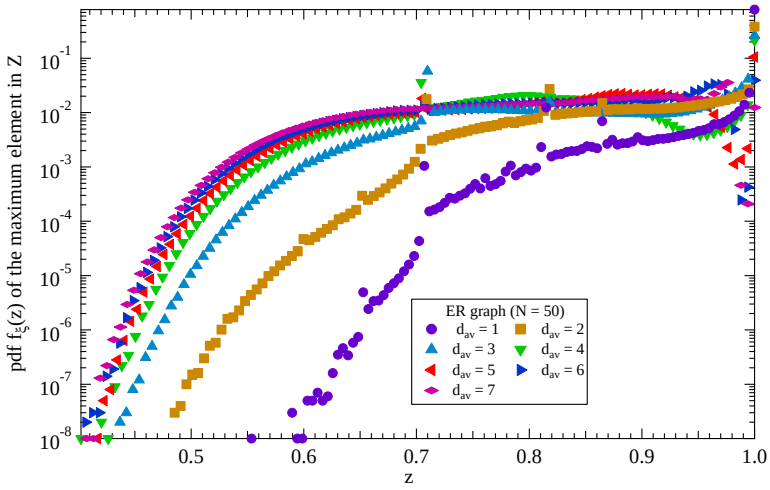


Figure 6.4: The probability density function $f_\zeta(z)$ of the maximum element in Z in ER random graphs $G_p(N)$ for $N = 50$ and various average degree $d_{av} = p(N-1)$, ranging from $d_{av} = 1$ up to $d_{av} = 7$. The y-axis is in log-scale.

onality condition (6.4) requires that the square of a row sum in Z equals one so that, for node j,

$$1 = \sum_{k=1}^{N} (z_k)_j^2 = \sum_{k=1}^{N-1} (z_k)_j^2 + \frac{1}{N}$$

implying that $\frac{1}{N} \leq \max_{1\leq k\leq N} (z_k)_j^2 \leq 1 - \frac{1}{N}$. Hence, in any connected graph, we find

that $\frac{1}{\sqrt{N}} \leq \xi_Z \leq \sqrt{1-\frac{1}{N}} < 1$ and, similarly, $-\sqrt{1-\frac{1}{N}} \leq \zeta_Z \leq -\frac{1}{\sqrt{N}}$.

6.4. DUAL FUNDAMENTAL WEIGHT VECTOR φ

In this section, we study, both numerically and analytically, the distribution of a random component in the dual fundamental weight vector φ, defined in (6.6). First, we note [161] that the sum s_{Z^2} of the elements of Z^2 is

$$s_{Z^2} = u^T Z^2 u = \omega^T \varphi$$

and with $\omega = \sqrt{N} e_N$, we have for a connected graph,

$$s_{Z^2} = \sqrt{N} \varphi_N$$

where $\varphi_N = \sum_{j=1}^{N} (z_N)_j$ is the N-th row sum of Z.

6.4.1. RANDOMLY CHOSEN COMPONENT OF THE DUAL FUNDAMENTAL WEIGHT VECTOR φ

As shown in [161], the vector ω is invariant with respect to a node relabeling transformation, but the dual fundamental weight vector φ is not, nor is s_{Z^2}. The consequence is that, by generating Erdős-Rényi random graphs, the node labeling is uniformly distributed so that the random variable $s_{Z^2} \overset{d}{=} \sqrt{N} \varphi_U$, where $U \in [1, N]$ is a discrete uniform random variable.

The expectation of a randomly chosen element φ_U is

$$E[\varphi_U] = \sum_{k=1}^{N} \varphi_k \Pr[U = k] = \frac{1}{N} \sum_{k=1}^{N} \varphi_k = \frac{1}{N} u^T \varphi$$

Since $u^T \varphi = u^T \omega = \sqrt{N}$ (see [161]), we find that

$$E[\varphi_U] = \frac{1}{\sqrt{N}} \tag{6.8}$$

The variance of φ_U, $\mathrm{Var}[\varphi_U] = E[\varphi_U^2] - (E[\varphi_U])^2$ follows, with $\sum_{k=1}^{N} \varphi_k^2 = N$ (see [161]) from

$$E[\varphi_U^2] = \sum_{k=1}^{N} \varphi_k^2 \Pr[U = k] = \frac{1}{N} \sum_{k=1}^{N} \varphi_k^2 = 1$$

so that

$$\mathrm{Var}[\varphi_U] = 1 - \frac{1}{N} \tag{6.9}$$

Extensive simulations on φ_U in Erdős-Rényi random graphs $G_p(N)$ are performed. We simulate ER random graphs for various N, where $N = 10, 20, 30, \cdots, 100$ and with the link density $p = 0.3$. For each N, we have simulated 10^8 ER random graphs that resulted in 10^8 realizations of φ_U. The probability density function $f_{\varphi_U}(z)$ for each N is plotted and fitted.

Next, we show that φ_U does not depend on the degree vector d for a regular graph. We start from

$$d^T\varphi = \sum_{k=1}^{N} d_k\varphi_k = N\sum_{k=1}^{N} d_k\varphi_k \Pr[U=k] = NE\left[d_U\varphi_U\right]$$

Thus, the correlation coefficient

$$\rho\left(d_U,\varphi_U\right) = \frac{1}{N}d^T\varphi - E[d_U]\,E\left[\varphi_U\right] = \frac{1}{N}d^T\varphi - \frac{2L}{N}\frac{1}{\sqrt{N}}$$

and

$$\rho\left(d_U,\varphi_U\right) = \frac{1}{N}\left(d^T\varphi - \frac{2L}{\sqrt{N}}\right)$$

The dependence or correlation between the degree vector d and the dual fundamental weight vector φ is zero provided $d^T\varphi = \frac{2L}{\sqrt{N}}$. In a regular graph, for example, $d = ru$, $\frac{2L}{\sqrt{N}} = r\sqrt{N}$ and $d^T\varphi = ru^T\varphi = ru^T\omega = r\sqrt{N}$, so that $\rho\left(d_U,\varphi_U\right) = 0$. Simulations hint that $\rho\left(d_U,\varphi_U\right) \approx 0$ for ER random graphs, too! Fig. 6.5 demonstrates that the probability density function $f_{\varphi_U}(z)$ is approximately an invariant with respect to the average degree d_{av} (and thus the link density p in $G_p(N)$).

6.4.2. THE PRODUCT OF A GAUSSIAN AND A SUPER-GAUSSIAN DISTRIBUTION

The probability density function $f_{\varphi_U}(z)$ is accurately fitted by the probability density function

$$f_X(z) = c\exp\left[-b(z-z_0)^2\right]\exp\left[-a(z-z_0)^4\right] \tag{6.10}$$

which is a product of a Gaussian and a super-Gaussian distribution. A random variable Y_m possesses a super-Gaussian distribution, defined by

$$f_{Y_m}(z) = A_m\exp\left[-a(z-z_0)^m\right]$$

where m is an even integer and $a > 0$ is a positive real number.

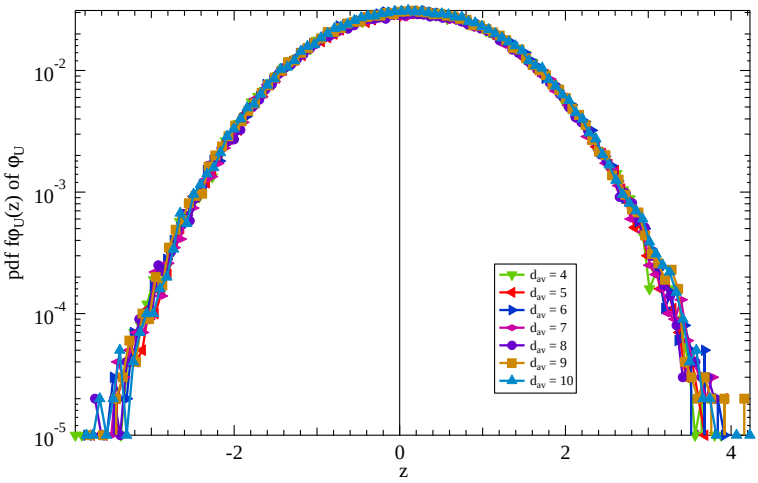


Figure 6.5: The probability density function $f_{\varphi_U}(z)$ of φ_U for connected ER random graphs $G_p(N)$ for $N = 50$ and vaious average degree d_{av}, ranging from $d_{av} = 4$ up to $d_{av} = 10$. The y-axis is in log-scale.

Next, we focus on determining the parameters a, b and c in (6.10). Since $\int_{-\infty}^{\infty} f_X(z)dz = 1$, with $z - z_0 = x$, we have

$$c\int_{-\infty}^{\infty} \exp[-bx^2 - ax^4]dz = 1$$

The integral, proved in [162],

$$\int_0^{\infty} \exp\left[-bu^2 - au^4\right] du = \frac{1}{4}\sqrt{\frac{b}{a}} e^{\frac{b^2}{8a}} K_{\frac{1}{4}}\left(\frac{b^2}{8a}\right)$$

and where $K_s(z)$ is the modified Bessel function of the Second Kind [163], determines c as

$$c = \frac{1}{2\int_0^{\infty} \exp\left[-bu^2 - au^4\right] du} = \sqrt{\frac{a}{b}} \frac{2e^{-\frac{b^2}{8a}}}{K_{\frac{1}{4}}\left(\frac{b^2}{8a}\right)} \tag{6.11}$$

Since $f_X(z)$ is a symmetric function around z_0, all odd centered moments around z_0, $E\left[(X - z_0)^k\right] = \int_{-\infty}^{\infty} (x - z_0)^k f_Z(x)\, dx$, are zero and, thus $E[X] = z_0$. Combination with (6.8) shows that $z_0 = \frac{1}{\sqrt{N}}$. We can compute the variance $\mathrm{Var}[X] = E\left[(X - z_0)^2\right]$ explicitly as

$$\mathrm{Var}[X] = \frac{1}{2b} h\left(\frac{y^2}{8}\right) \tag{6.12}$$

with

$$h(t) = 2t\left(\frac{K_{\frac{3}{4}}(t)}{K_{\frac{1}{4}}(t)} + \frac{K_{\frac{5}{4}}(t)}{K_{\frac{1}{4}}(t)} - 2\right) - 1$$

where $y^2 = \frac{b^2}{a}$. Further, Var[X] is increasing with y from 0 (for $y = 0$) to $\frac{1}{2b}$ (when $y \to \infty$). Using (6.9) yields

$$b = \frac{h\left(\frac{y^2}{8}\right)}{2\left(1 - \frac{1}{N}\right)} \tag{6.13}$$

while $y^2 = \frac{b^2}{a}$ then leads to

$$a = \frac{h^2\left(\frac{y^2}{8}\right)}{4y^2\left(1 - \frac{1}{N}\right)^2} \tag{6.14}$$

Hence, (6.13) and (6.14) indicate that b increases with y towards $\frac{1}{2\left(1-\frac{1}{N}\right)}$, while a decreases with y towards 0.

6.4.3. FITTING RESULT

Fig. 6.6 shows the natural logarithm of the probability density function $f\varphi_U(z)$ for φ_U from simulations, fitted by the function (6.10). As observed from Fig. 6.6, the simulations agree astonishingly well with (6.10) for all N simulated in this chapter.

Fig. 6.7 shows that the parameter $y^2 = \frac{b^2}{a}$ is approximately linear in N,

$$y^2 = 0.5N - 3.85 \tag{6.15}$$

Substituting the linear function (6.15) into (6.14) and (6.13) determines a and b analytically. As shown in Figs. 6.8 and 6.9, a and b (red curve, theory from (6.14) and (6.13) with (6.15)) agree well with simulations of φ_U (black dots), after fitting a and b from (6.10). Fig. 6.10 shows c from (6.11) and from fitting function (6.10) for $f_{\varphi_U}(z)$ for each N. Fig. 6.11 presents z_0 from (6.8) and from the fitting function (6.10).

As shown in Fig. 6.8-6.11, the fitting parameters a, b, c, z_0 in (6.10) from simulations agree well with equations (6.14), (6.13), (6.11), (6.8), respectively. Thus, our simulations lead us to believe that the distribution of the components of the dual fundamental weight vector φ in Erdős-Rényi random graphs is given by (6.10), which is the product of a Gaussian and a super-Gaussian. Fig. 6.8 and (6.14) (with (6.15)) show that a tends as $O(1/N)$ to zero with N, implying that, for large N, the super-Gaussian disappears and the expected Gaussian behavior (from random matrix theory) appears. The parameter a in (6.10) constraints the Gaussian behavior, which is likely due to the orthogonality conditions (6.2) and (6.4) that create dependence among the eigenvector components. Indeed, the larger N, the less the or-

thogonality conditions are confining, which suggest that a would decrease inversely proportional to N, precisely as observed in Fig. 6.8.

6.4.4. VERY SMALL SIZES OF N

We observe that when $N < 8$ (obtained at the point $y^2 < 0$ in (6.15)), the simulation result is better fitted by a Gaussian distribution, instead of the product of a Gaussian and a super-Gaussian.

As shown in Fig. 6.12, the product of a Gaussian and super-Gaussian distribution does not precisely fit the simulations at the tail. When N is decreased to 6 in Fig. 6.13, the simulation is fitted by a Gaussian distribution.

6.5. CHAPTER CONCLUSION

We have studied the eigenvector matrix Z of the Laplacian matrix Q for a graph G with the aim to understand how properties of Z contain information about the structure of G. We find that the sum s_Z of all the elements in Z increases with the size of the graph as $O(\sqrt{N})$. The higher the average degree in a graph, the lower the number of zeros in the eigenvector matrix. Moreover, the distribution of the maximum element in the eigenvector matrix is the same as the distribution of the minimum element.

The row sum of the eigenvector matrix Z of the Laplacian Q, coined the dual fundamental weight φ, in Erdős-Rényi random graphs follows closely the product of a Gaussian and a super-Gaussian distribution.

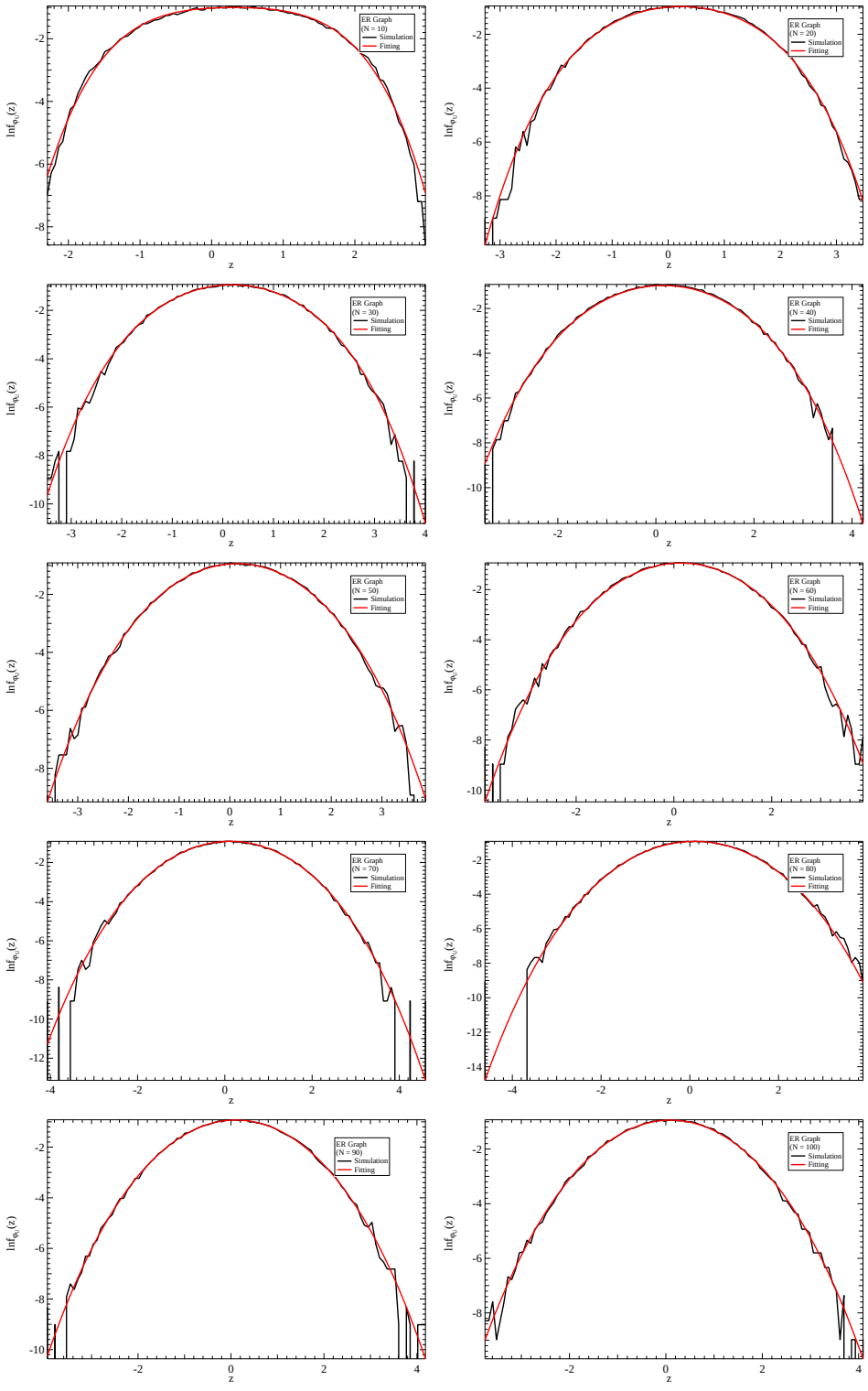


Figure 6.6: Natural logarithm $\ln(f\varphi_U(z))$ of the probability density function $f_{\varphi_U}(z)$ for ER graphs with $p = 0.3$ and various N, ranging from $N = 10$ to $N = 100$.

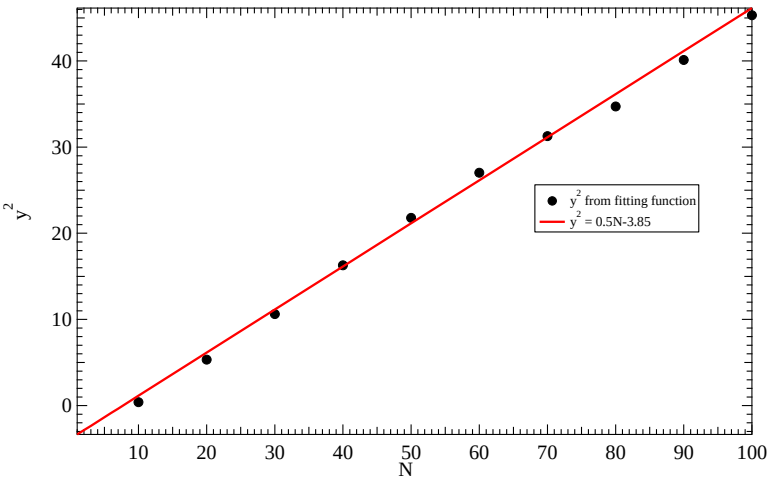


Figure 6.7: Fitting parameter $y^2 = \frac{b}{a}$ as a function of N in ER graphs.

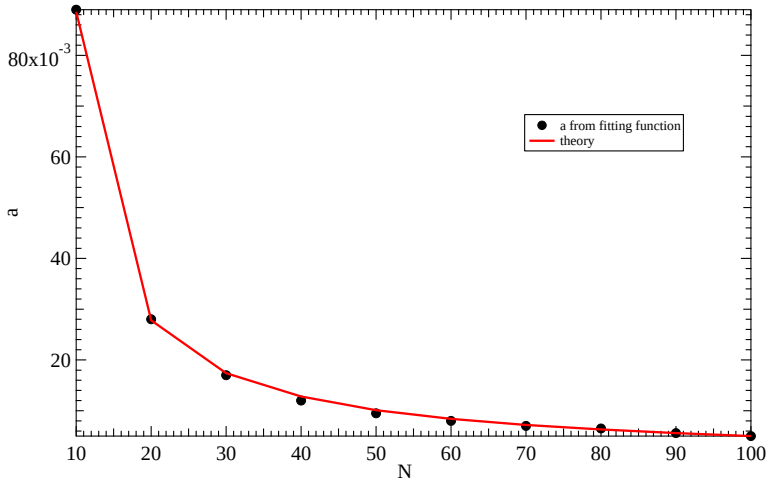


Figure 6.8: Fitting parameter a as a function of N in ER graphs.

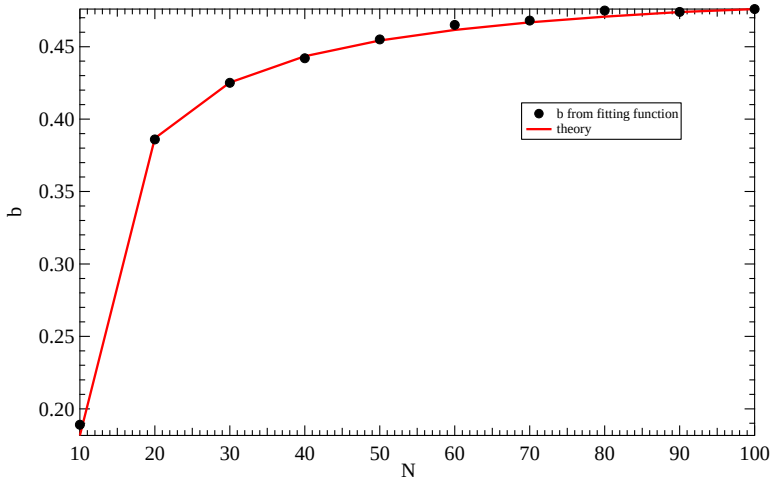


Figure 6.9: Fitting parameter b as a function of N in ER graphs.

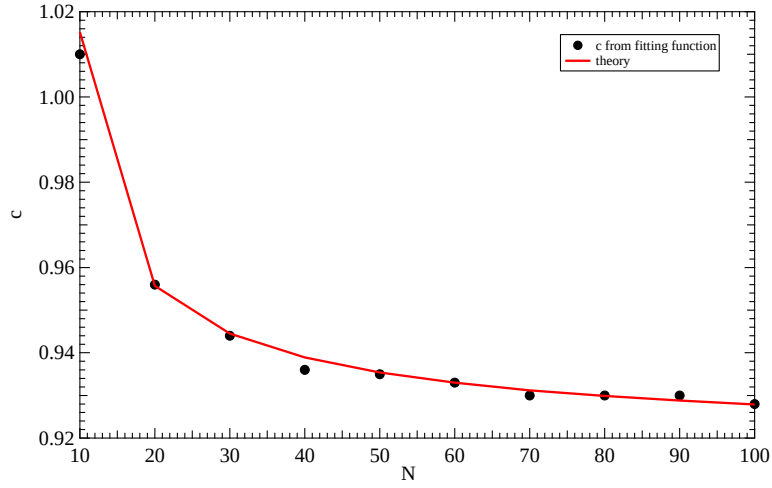


Figure 6.10: Fitting parameter c as a function of N in ER graphs.

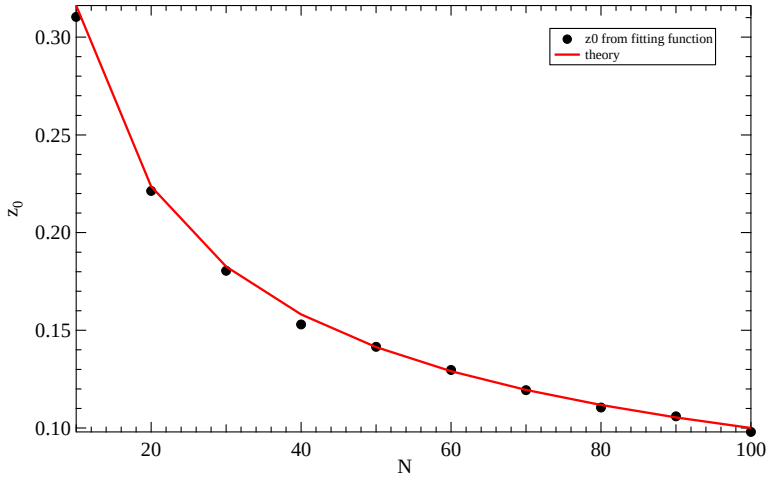


Figure 6.11: Fitting parameter z_0 as a function of N in ER graphs.

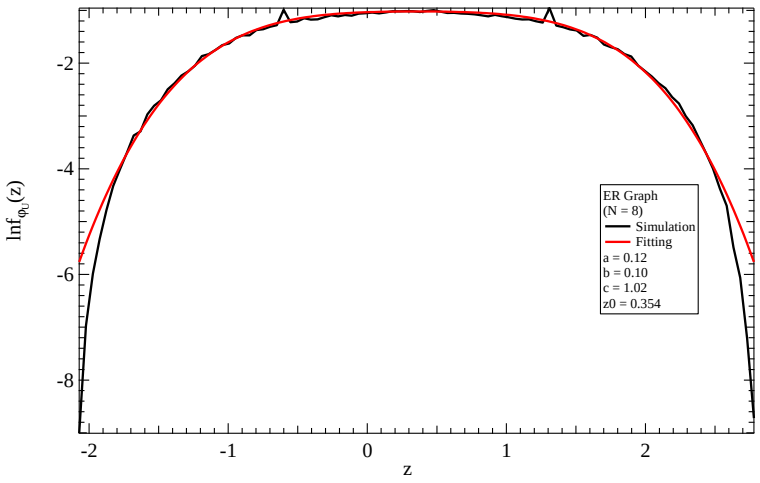


Figure 6.12: Natural logarithm $\ln(f\varphi_U(z))$ of the probability density function $f_{\varphi U}(Z)$ for 10^8 ER graphs with $p = 2\log(N)/N$ (to make sure the graph is connected) and $N = 8$.

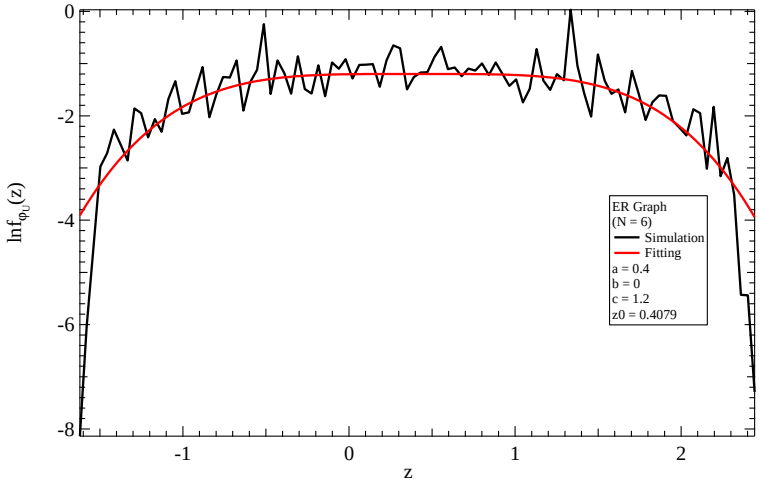


Figure 6.13: Natural logarithm $\ln(f\varphi_U(z))$ of the probability density function $f_{\varphi_U}(Z)$ for 10^8 ER graphs with $p = 2\log(N)/N$ (to make sure the graph is connected) and $N = 6$.

Part III: Robustness of interdependent networks

7

MODELLING REGION-BASED INTERCONNECTION FOR INTERDEPENDENT NETWORKS

7.1. INTRODUCTION

In the real world, most networks are interdependent. For example, power networks depend on communication networks, where each node in a communication network controls one or more nodes in a power network, while each communication node needs power to function [164]. Most infrastructures are interdependent networks, such as transportation networks, communications and energy supply networks. An interdependent network is a network consisting of different types of networks that interact with each other via interconnected links [165].

In interdependent networks, a cascade of failures leads to the first-order (discontinuous) percolation transition whereas a second-order (continuous) phase transition characterizes the collapse of a single network [35, 166]. Some types of interdependent networks also feature a structural transition [167] between distinguishable and non-distinguishable network components. The exact transition threshold for

such a structural transition is determined in [168]. Most previous studies are restricted to a one-to-one interdependency between networks, where one-to-one interdependency means that one node in one network connects to one and only one node in the other network and vice versa. Boccaletti *et al.* [166] introduce models that enable nodes in one network connect to multiple nodes in the other network, with a given degree sequence for interconnections. Moreover, the location of the nodes is not considered when designing the interconnection between interdependent networks, although connecting geographically close nodes is less costly than connecting those that are far away from each other.

We propose two topologies, the random geometric graph and the relative neighbourhood graph, that incorporate the location of nodes for the design of interconnection in interdependent networks. The advantages of the models are that (i) the interdependency is generalized from one-to-one to one-to-many interconnections; (ii) the sizes of the interdependent networks are not necessarily equal.

We derive the average number of links for the two topologies which enables the comparison between simulations performed on them. For the two topologies, we investigate the impact of the interconnection structure on the robustness of the network under node failures. The size of the largest mutually connected component (the number of functioning nodes) is employed as a robustness metric. In addition, we propose the derivative of the largest mutually connected component with respect to the fraction of failed nodes as a new robustness metric. The proposed robustness metric quantifies the damage on the whole network triggered by a small fraction of non-functioning nodes.

The paper is organized as follows. Section 7.2 illustrates two interconnection topologies that incorporate the location of nodes. Section 7.3 presents the cascading failures in interdependent networks. The simulation results are presented in Section 7.4 and Section 7.5 concludes the paper.

7.2. REGION-BASED INTERDEPENDENCY

Consider an interdependent graph $G(N, L)$ with N nodes and L links consisting of two graphs G_1 and G_2. The adjacency matrix A of G can be written as

$$A = \begin{bmatrix} (A_1)_{n\times n} & B_{n\times m} \\ \left(B^T\right)_{m\times n} & (A_2)_{m\times m} \end{bmatrix} \tag{7.1}$$

where A_1 is the $n \times n$ adjacency matrix of the graph G_1 with n nodes, A_2 is the $m \times m$ adjacency matrix of the graph G_2 with m nodes and B is the $n \times m$ interconnection matrix connecting G_1 and G_2. The total number of nodes in G is $N = n + m$. The interaction between networks G_1 and G_2 completely relies on the interconnection matrix B. The design of B is, therefore, crucial for the interdependent networks to function properly as a whole.

In this paper, we propose two topologies for the interconnection matrix B incorporating the geographical location of nodes. Associating each node with a coordinate, we analyse the interconnection matrix B with elements b_{ij} in the following two ways: $b_{ij} = 1$ if

1. *random geometric graph* [169]: the Euclidean distance d_{ij} between node i and node j is smaller than a given threshold r;

2. *relative neighbourhood graph* [170]: there is no third node in the intersection region of two circles with centres at nodes i and j with the same radius equal to their Euclidean distance d_{ij}.

Figure 7.1 shows the two topologies of the interconnection matrix B.

7.2.1. RANDOM GEOMETRIC GRAPH

A random geometric graph, denoted as $G_{p_{ij}}(N)$, consists of N nodes and two nodes i and j are connected by a link with probability p_{ij}. Consider N independent and identically distributed nodes in a two-dimensional square with size Z. Any square with size Z can be normalized [171] to a unit square ($Z = 1$) without changing the probability p_{ij}. For simplicity, we consider a unit square with size $Z = 1$. Let (x_i, y_i) and (x_j, y_j) be the coordinates for nodes i and j as illustrated in Figure 7.2. Let $r \geq 0$ be a non-negative and real number which is referred to as the radius of a

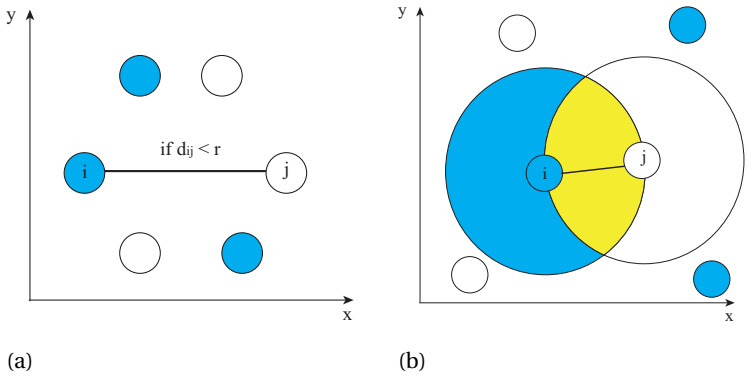


Figure 7.1: Two topologies for B: (a) random geometric graph; (b) relative neighbourhood graph: since there is no third node in the intersection region (marked as yellow), nodes i and j are connected. Nodes from G_1 are represented with filled circles, whereas nodes from G_2 are represented with unfilled circles.

node. The probability $p_{ij}(r) = \Pr[d_{ij} \leq r]$ is the probability that the Euclidean distance $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ between two uniformly distributed nodes i and j is less than or equal to the radius r. The maximum Euclidean distance between two nodes in a two-dimensional square with size $Z = 1$ is $\sqrt{2}$. When $r \geq \sqrt{2}$, the probability for nodes i and j being connected is $p_{ij} = 1$ and thus, the graph $G_{p_{ij}}(N)$ is a complete graph K_N.

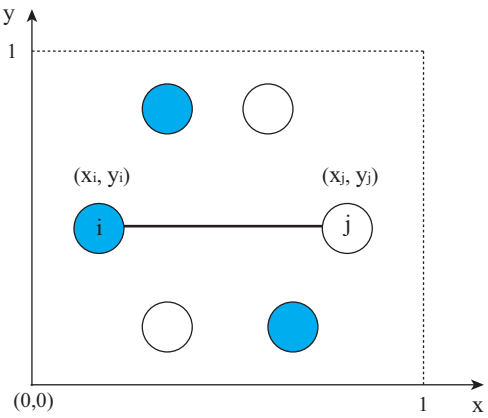


Figure 7.2: Node coordinate

In subsection 7.2.1, we prove a theorem for p_{ij} in a general random geometric graph in a two-dimensional square with size $Z = 1$.

PROBABILITY p_{ij} OF HAVING A LINK BETWEEN NODES i AND j

Theorem 8. *The probability $p_{ij}(r)$ that there is a link l_{ij} between nodes i and j in a random geometric graph in a two-dimensional unit square is*

$$p_{ij}(r) = \begin{cases} \pi r^2 - \frac{8}{3} r^3 + \frac{1}{2} r^4 & 0 \le r \le 1 \\ \frac{1}{6}\left[-3r^4 + (16r^2+8)\sqrt{r^2-1} + 12r^2\left(\arctan\left(\frac{2-r^2}{2\sqrt{r^2-1}}\right) - 1\right) + 2\right] & 1 < r \le \sqrt{2} \\ 1 & \sqrt{2} < r \end{cases}$$

Proof. The probability $p_{ij}(r)$ that there is a link l_{ij} between nodes i and j in a square with size $Z = 1$ is

$$p_{ij}(r) = \Pr\left[(x_i - x_j)^2 + (y_i - y_j)^2 \le r^2\right]$$

Let $Z_1 = |X_1 - X_2|$ and $Z_2 = |Y_1 - Y_2|$ be random variables. The probability distribution function for Z_1 is, when $0 \le z_1 \le 1$,

$$F(z_1) = \Pr[-z_1 \le X_1 - X_2 \le z_1]$$

Since X_1 and X_2 are independent uniform random variables, we obtain

$$\begin{aligned} \Pr[X_1 - X_2 \le z_1] &= \int_0^{1-z_1} \int_0^{x_2+z_1} dx_1 dx_2 + \int_{1-z_1}^{1} \int_0^{1} dx_1 dx_2 \\ &= \frac{1}{2}\left(1 - z_1^2\right) + z_1 \end{aligned}$$

Analogously,

$$\Pr[X_1 - X_2 \le -z_1] = \frac{1}{2}(z_1 - 1)^2$$

With $F(z_1) = \Pr[X_1 - X_2 \le z_1] - \Pr[X_1 - X_2 \le -z_1]$, we arrive at

$$F(z_1) = -z_1^2 + 2z_1$$

The probability density function $f(z_1) = F'(z_1)$ follows, when $0 \le z_1 \le 1$,

$$f(z_1) = 2(1 - z_1)$$

Since Z_1 and Z_2 are independent and identically distributed, we have

$$\Pr\left[(x_i - x_j)^2 + (y_i - y_j)^2 \le r^2\right] = \int\int_{z_1{}^2 + z_2{}^2 \le r^2} f(z_1) f(z_2) dz_1 dz_2$$

For $0 \le r \le 1$, we have, after transformation to polar coordinates,

$$\begin{aligned}\Pr\left[(x_i - x_j)^2 + (y_i - y_j)^2 \le r^2\right] &= \int_0^r \int_0^{\sqrt{r^2 - z_2^2}} f(z_1) f(z_2) dz_1 dz_2 \\ &= \pi r^2 - \frac{8}{3} r^3 + \frac{1}{2} r^4 \end{aligned} \tag{7.2}$$

Similarly, we find, for $1 < r \le \sqrt{2}$,

$$\begin{aligned}\Pr\left[(x_i - x_j)^2 + (y_i - y_j)^2 \le r^2\right] &= \int_0^1 \int_0^{\sqrt{r^2-1}} f(z_1) f(z_2) dz_1 dz_2 + \int_{\sqrt{r^2-1}}^1 \int_0^{\sqrt{r^2 - z_1^2}} f(z_1) f(z_2) dz_1 dz_2 \\ &= \frac{1}{3} + 2r^2 \left(\arctan\left(\frac{2 - r^2}{2\sqrt{r^2 - 1}} \right) - 1 \right) + \frac{8r^2 + 4}{3} \sqrt{r^2 - 1} - \frac{1}{2} r^4 \end{aligned} \tag{7.3}$$

For $r > \sqrt{2}$, the distance between two nodes in a unit square is always less than or equal to $\sqrt{2}$. Hence, the probability $p_{ij}(r)$ is always 1. Combining (7.2) and (7.3) establishes Theorem 8. □

Figure 7.3 shows the probability p_{ij} as a function of the radius r in a random geometric graph $G_{p_{ij}}(N)$ with $N = 10^4$ nodes. The simulation shows an excellent agreement with Theorem 8. From Theorem 8, the average number of links for a

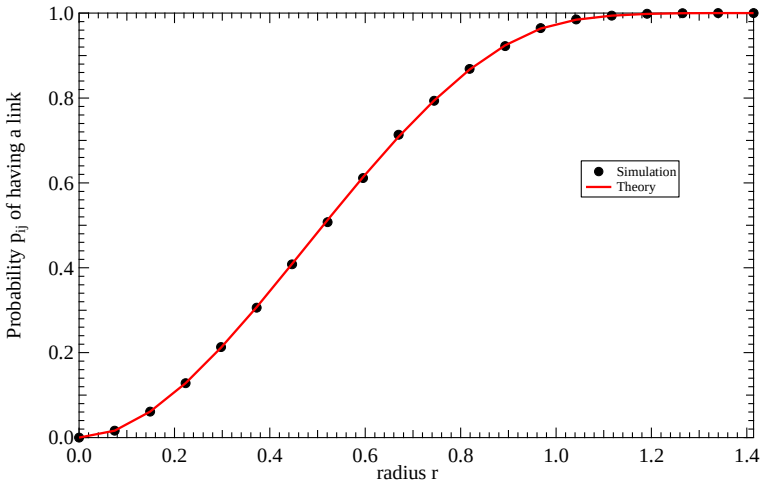
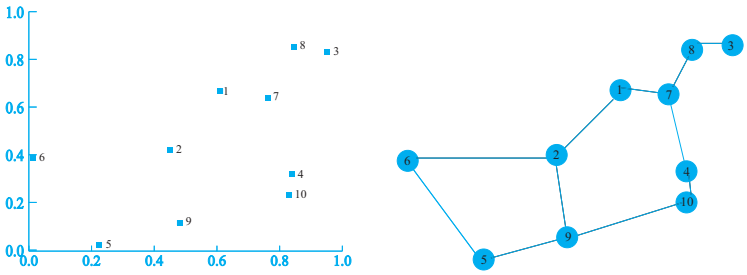


Figure 7.3: The probability $p_{ij}(r)$ that nodes i and j are connected as a function of the radius r in a random geometric graph with $N = 10^4$ nodes.

random geometric graph with N nodes is $E[L] = \binom{N}{2} p_{ij}(r)$.

7.2.2. RELATIVE NEIGHBOURHOOD GRAPH

A relative neighbourhood graph, denoted as $RNG(N)$, consists of N nodes and two nodes i and j are connected if $d_{ij} \leq \max(d_{ik}, d_{jk})$ for all the other nodes $k = 1,2,\ldots,N$, $k \neq i, j$. Figure 7.4 shows a set of N nodes in a two-dimensional square with size $Z = 1$ and its relative neighbourhood graph. In subsection 7.2.2, we prove a theorem for the lower bound of the probability p_{ij} of nodes i and j being connected in a general relative neighbourhood graph.



(a) (b)

Figure 7.4: An example of (a) a set of N nodes and its (b) relative neighbourhood graph.

7

PROBABILITY p_{ij} OF HAVING A LINK BETWEEN NODES i AND j

Theorem 9. *The probability p_{ij} that for a relative neighbourhood graph there is a link l_{ij} between nodes i and j in a two-dimensional square with size $Z = 1$ is lower bounded by*

$$p_{ij} \geq \frac{\pi c N + 1}{c^2 N(N-1)} - \frac{2\sqrt{\pi}\Gamma(N-1)}{c^{\frac{3}{2}}\Gamma\left(N+\frac{1}{2}\right)} \tag{7.4}$$

where $c = \left(\frac{2\pi}{3} - \frac{\sqrt{3}}{2}\right)$ and $\Gamma(x)$ is the gamma function.

Proof. Given a pair of nodes i and j uniformly distributed in the square with size $Z = 1$, let A be the random variable for the area of the intersection region (marked as yellow in Fig. 7.1(b)) of two circles centred at nodes i and j and with d_{ij} as the radius. For a two-dimensional square with size $Z = 1$, the area of the square is 1. The probability p_{ij} that nodes i and j being connected equals the probability that all the other $N-2$ nodes are not in the intersection region A:

$$p_{ij} = (1 - A)^{N-2}$$

Using the law of total probability [33], we have

$$p_{ij} = \int_0^1 (1-x)^{N-2} f_A(x) dx \tag{7.5}$$

where $f_A(x)$ is the probability density function of A. The probability distribution function for the variable A is

$$F_A(x) = \Pr[A \leq x]$$

Let D be the random variable of the distance between two nodes. The area [172] of the intersection of two circles can be computed by $D^2 c$, where $c = \left(\frac{2\pi}{3} - \frac{\sqrt{3}}{2}\right)$. When the intersection is completely in the two-dimensional unit square, it holds that $A = D^2 c$. When the intersection is partially in the unit square, we have, for $\varepsilon > 0$, that $A + \varepsilon = D^2 c$ and, hence,

$$\begin{aligned} F_A(x) &= \Pr\left[D^2 c - \varepsilon \leq x\right] \\ &\geq \Pr\left[D^2 c \leq x\right] \end{aligned}$$

Applying $D^2 = (x_i - x_j)^2 + (y_i - y_j)^2$ and $r^2 = \frac{x}{c} < 1$ in (7.2) yields,

$$\Pr\left[D^2 c \leq x\right] = \frac{\pi x}{c} - \frac{8}{3}\left(\frac{x}{c}\right)^{\frac{3}{2}} + \frac{1}{2}\left(\frac{x}{c}\right)^2$$

The probability distribution function is lower bounded by

$$F_A(x) \geq \frac{\pi x}{c} - \frac{8}{3}\left(\frac{x}{c}\right)^{\frac{3}{2}} + \frac{1}{2}\left(\frac{x}{c}\right)^2$$

from which

$$f_A(x) \geq \frac{\pi}{c} - 4\left(\frac{x}{c^3}\right)^{\frac{1}{2}} + \frac{x}{c^2}$$

Thus, we have for (7.5)

$$p_{ij} \geq \int_0^1 (1-x)^{N-2} \left(\frac{\pi}{c} - 4\left(\frac{x}{c^3}\right)^{\frac{1}{2}} + \frac{x}{c^2}\right) dx \tag{7.6}$$

Using the Beta function $B(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ in (7.6), we establish Theorem 9. □

It has been shown [170] that the relative neighbourhood graph is a superset of the minimum spanning tree. The number L of links in the relative neighbourhood graph with N nodes is bounded [170] by

$$N - 1 \leq L \leq 3N - 6 \tag{7.7}$$

Hence, the link density $p = \frac{L}{\binom{N}{2}}$ for a relative neighbourhood graph is bounded by $\frac{2}{N} \leq p \leq \frac{6(N-2)}{N(N-1)}$ which shows that the relative neighbourhood graph is a sparse graph: the larger the size N of the graph, the sparser the graph is. From Theorem 9, we deduce the lower bound for the average number $E[L]$ of links

$$E[L] \geq \binom{N}{2} p_{ij} \tag{7.8}$$

A different lower bound for $E[L]$ is presented in [173]

$$E[L] \geq 0.689N \tag{7.9}$$

Figure 7.5 shows the average number of links $E[L]$ for $RNG(N)$ with N ranging from 50 to 200. Figure 7.5 shows that our bound (7.8) is close to the simulations and outperforms bound (7.9).

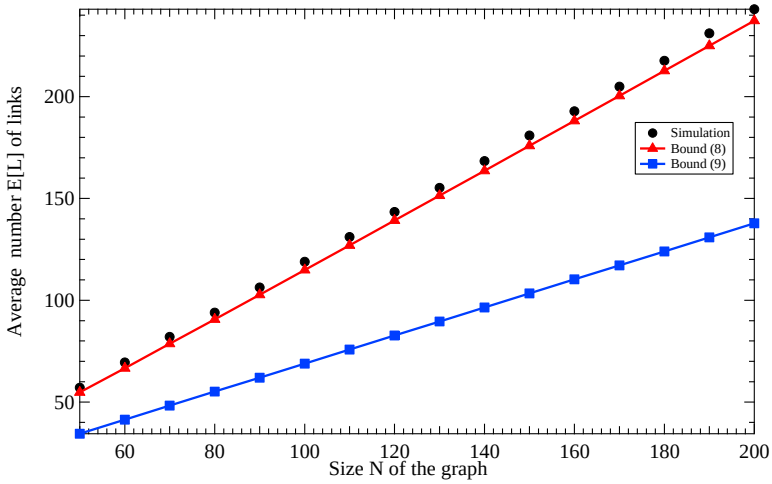


Figure 7.5: Number of links for $RNG(N)$ with N ranging from 50 to 200.

7.3. CASCADING FAILURES IN INTERDEPENDENT NETWORKS

When nodes in one network fail, the interconnection structure between two networks causes dependent nodes in the other network also to fail. This may happen recursively and may invoke a cascading failure until no more nodes fail. In this section, we investigate the impact of interconnection topologies on the robustness

of interdependent networks against cascading failures. The robustness is quantified by (i) Largest Mutually Connected Component (LMCC); (ii) derivative of the largest mutually connected component with respect to the fraction of removed nodes.

7.3.1. LARGEST MUTUALLY CONNECTED COMPONENT

Different from the models [35, 174, 175] where a node from one network depends on one and only one node from the other network (one-to-one interconnection), we generalize the interconnection pattern to one-to-many: a node might depend on zero or one or more than one node depending on the distance to other nodes.

In our model, we assume a node n_1 in network G_1 to be functional if (i) its interdependent nodes in network G_2 are functioning; (ii) the node belongs to the giant component of the functional nodes in network G_1. Since a node n_1 in G_1 may have more than one support node in G_2, we assume two scenarios for n_1 being supported by nodes in G_2: (i) *at least one* of the supported nodes in G_2 is functioning; (ii) *all* of its supported nodes in G_2 are functioning. The same assumptions are applied to the nodes in network G_2.

A random removal of a fraction $1-q$ of nodes in network G_1, on one hand, isolates nodes in network G_1 and on the other hand causes nodes in network G_2 to fail because of removed interconnected nodes in G_1. The failed nodes in network G_2 isolate nodes from the giant component in networks G_2. The isolated nodes in G_2 further introduce failures in G_1 and so on. The cascading failures continue until no more nodes are failed. The remaining set of functional nodes is referred to the *largest mutually connected component (LMCC)*. We assume, without loss of generality, that the fraction $1-q$ of nodes is removed from graph G_1.

ALGORITHM DESCRIPTION

The metacode for computing the largest mutually connected component is given in Algorithm 1. The main algorithm starts at line 3 where n is the number of realizations of G. Lines 4 to 16 generate an interdependent graph G consisting of either two Erdős-Rényi (ER) graphs or two Barabási-Albert (BA) graphs. The interconnection topology is either the random geometric graph (RGG in line 12) or the relative neighbourhood graph (RNG in line 14). From line 17 to line 22, we com-

pute the largest mutually connected component after cascading failures triggered by $1-q$ removals. Lines 23 and 25 average the largest mutually connected component over n instances of G. The metacode for function CASCADING (line 20) and COMPONENT (line 21) is given in the Appendix D.1.

We elaborate on two special values of $1-q$, i.e., 0 and $\frac{N_1}{N}$. For $1-q=0$, we assume LMCC = 1. We encounter a special scenario that there exists nodes without supporting nodes before any removals, as shown in Figure 7.6, due to their location being far away from nodes in the other network. We assume such nodes are alive until they are isolated from their own network. When $1-q=\frac{N_1}{N}$, the nodes in graph G_1 are completely removed. Nodes in G_2 have no supporting nodes from G_1 and thus also fail. Hence, there is no largest mutually connected component and $LMCC=0$.

Figure 7.7 exemplifies Algorithm 1 when G_1 and G_2 are complete graphs and the interconnection matrix is $B=J$ where J is the all one matrix representing all-to-all interconnections. We assume that a node is alive if at least one of its supporting nodes is alive. Figure 7.7 shows that when $1-q=0$, the interdependent network is fully connected and LMCC = 1. With the increase of $1-q$ removals, LMCC decreases linearly with $1-q$. The slope of the line is -1. When $1-q=\frac{N_1}{N}$ (0.5 in Figure 7.7), the nodes in graph G_1 are completely removed and $LMCC=0$.

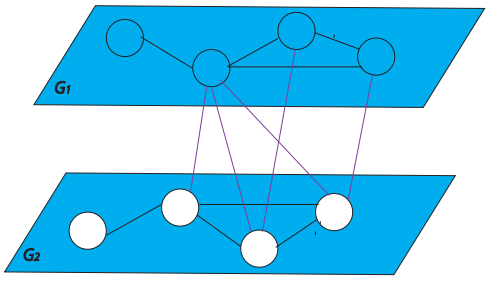


Figure 7.6: An interdependent network with nodes having no interconnected nodes

7.3.2. DERIVATIVE FOR THE LARGEST MUTUALLY CONNECTED COMPONENT

In the real world, a network that completely collapses is a disaster for network providers. To avoid the disaster, understanding the impact of the failure of a relatively small fraction, e.g. 10%, of nodes is significant for network providers. We theoretically approach the robustness of interdependent networks under a small

Algorithm 1 AverageLMCC

Input: Sizes N_1 and N_2 for graphs G_1 and G_2, respectively; The parameter graph specifies G_1 and G_2 to be ER graphs with link density p or BA graphs with m; The parameter interconnection specifies B to be RGG with radius r or RNG.
Output: Average of the largest mutually connected component (LMCC) over n graph instances.
for i=1 **to** n **do**
  if graph = ER **then**
    $G_1 \leftarrow \text{ER}(N_1, p)$ {generate an ER graph where nodes are connected with probability p }
    $G_2 \leftarrow \text{ER}(N_2, p)$
  else if graph = BA **then**
    $G_1 \leftarrow \text{BA}(N_1, m)$ {generate a BA graph where a new node with m links preferentially connects to high degree nodes}
    $G_2 \leftarrow \text{BA}(N_2, m)$
  end if
  if interconnection = RGG **then**
    $B \leftarrow \text{RGG}(N_1, N_2, r)$ {$N_1 \times N_2$ interconnection matrix where $B_{ij} = 1$ if $d_{ij} < r$ }
  else if interconnection = RNG **then**
    $B \leftarrow \text{RNG}(N_1, N_2)$ {$N_1 \times N_2$ interconnection matrix where $B_{ij} = 1$ if $d_{ij} \leq \max(d_{ik}, d_{jk})$ for all $k = 1, 2, \ldots, N$, $k \neq i, j$}
  end if
  $G \leftarrow \begin{bmatrix} G_1 & B \\ B^T & G_2 \end{bmatrix}$
  $\mathcal{N}_1 \leftarrow$ node labels of G_1 in G
  $\mathcal{N}_2 \leftarrow$ node labels of G_2 in G
  for $1 - q = 0$ **to** $\frac{N_1}{N}$ **step** 0.01 **do**
    $endGraph \leftarrow \text{CASCADING}(G, 1 - q, \mathcal{N}_1, \mathcal{N}_2)$
    $\text{T}_{1-q} \leftarrow \left|\text{COMPONENT}(endGraph, \mathcal{N}_1, \mathcal{N}_2)\right|$
  end for
  LMCC$[i] \leftarrow T$
end for
return mean(LMCC)

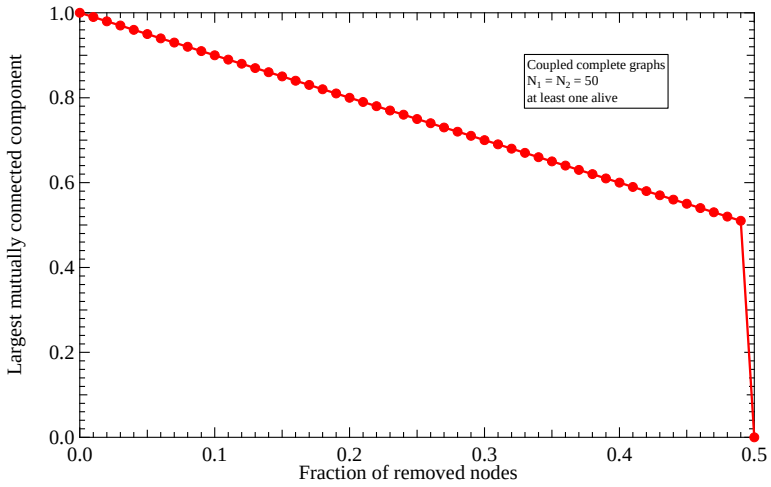


Figure 7.7: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are complete graphs and the interconnection matrix is $B = J$.

fraction of failures by investigating the derivative of the largest mutually connected component close to $1 - q = 0$. We suggest that this derivative can be used as a robustness measure of a network indicating the extent of damage on networks when a small fraction of nodes initially fails. The smaller the absolute derivative is, the higher robustness the network exhibits.

Starting from the derivative in a single network in subsection 7.3.2, we move step by step towards the derivative in interdependent networks with one-to-many interconnection in Subsection 7.3.2.

DERIVATIVE OF THE LARGEST CONNECTED COMPONENT FOR A SINGLE NETWORK

Given the probability generating function $\varphi_D(z)$ of the degree D of an arbitrary node, the probability generating function $\varphi_{(D_{l^+}-1)}$ of the degree of an end node l^+ reached by following an arbitrarily chosen link l is $\frac{\varphi'_D(z)}{\varphi'_D(1)}$, see [33]. Let $\varphi_{C_{l^+}}(z)$ be the generating function of the size C_{l^+} of components that are reached by following a random link l towards one of its end nodes l^+. If we choose a random node n in G and let $n = l^-$, we reach a component with generation function $\varphi_{C_n}(z)$ by following the link l towards the other end node l^+. If a node in the graph is occupied uniformly at random with probability q, the probability generating functions $\varphi_{C_{l^+}}(z)$

and $\varphi_{C_n}(z)$ follow [33]

$$\varphi_{C_{l^+}}(z) = 1 - q + qz\varphi_{(D_{l^+}-1)}\left[\varphi_{C_{l^+}}(z)\right]$$

$$\varphi_{C_n}(z) = 1 - q + qz\varphi_D\left[\varphi_{C_{l^+}}(z)\right]$$

Let S be the fraction of nodes in the largest connected component. Since $\varphi_{C_n}(z)$ generates the probability distribution of C_n excluding the giant component and with $\varphi_{C_n}(1) = 1$, we have that [33]

$$S = 1 - \varphi_{C_n}(1) = q - q\varphi_D\left[\varphi_{C_{l^+}}(1)\right]$$

where

$$\varphi_{C_{l^+}}(1) = 1 - q + q\varphi_{(D_{l^+}-1)}\left[\varphi_{C_{l^+}}(1)\right] \tag{7.10}$$

The derivative of the largest connected component S with respect to q is

$$\frac{dS}{dq} = 1 - \varphi_D(u) - q\varphi'_D(u)u'$$

where $u = \varphi_{C_{l^+}}(1)$. The derivative of (7.10) follows

$$u' = \varphi_{(D_{l^+}-1)}(u) + q\varphi'_{(D_{l^+}-1)}(u)u' - 1$$

Combining $\frac{u+q-1}{q} = \varphi_{(D_{l^+}-1)}(u) = \frac{\varphi'_D(z)}{\varphi'_D(1)}$ and $\varphi'_D(1) = E[D]$, we arrive at

$$\frac{dS}{dq} = \frac{S}{q} - \frac{E[D](u-1)\left(u-1+q\right)}{q\left(1 - q\varphi'_{(D_{l^+}-1)}(u)\right)} \tag{7.11}$$

When graph G is a large ER random graph, there holds to a good approximation [33, p. 39] that $\varphi_D(z) = \varphi_{(D_{l^+}-1)}(z) = e^{E[D](z-1)}$. In that case, the derivative $\frac{dS}{dq}$ in (7.11) can be simplified, with $u = 1 - S$, to

$$\frac{dS}{dq} = \frac{S}{q\left(1 - E[D]\left(q - S\right)\right)}$$

Figure 7.8 shows the straight line $y = -\frac{dS}{dq}\Big|_{1-q=\frac{1}{N}}\left(1-q\right) + 1$ and simulations of the largest mutually connected component. The straight line with slope $-\frac{dS}{dq}\Big|_{1-q=\frac{1}{N}}$ shows a good estimation for the largest mutually connected component when a small fraction $1 - q$ of nodes is removed.

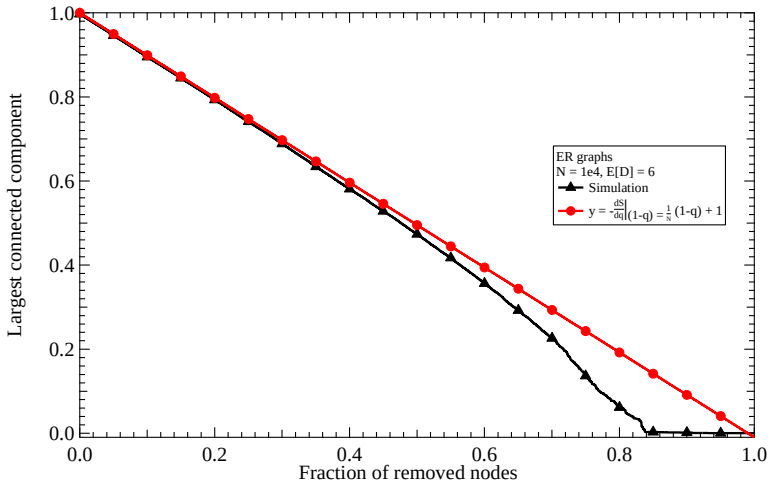


Figure 7.8: Largest connected component as a function of the fraction of removed nodes in Erdős-Rényi graphs $G_p(N)$.

DERIVATIVE FOR INTERDEPENDENT NETWORKS WITH ONE-TO-ONE INTERCONNECTION

Let $u_A = \varphi_{C_{l^+}}(1)$ for graph G_1 and $u_B = \varphi_{C_{l^+}}(1)$ for graph G_2. For interdependent networks with one-to-one interconnection, we have

$$u_A = \varphi_{(D_{l^+}-1)}(1 - q(1-u_B)(1-u_A)) \tag{7.12}$$

Analogously,

$$u_B = \varphi_{(D_{l^+}-1)}(1 - q(1-u_A))(1-u_B))$$

A randomly chosen node in G_1 belongs to the largest mutually connected component if (i) the node is occupied with probability q; (ii) the node with probability $1 - \varphi_{C_{G_1}}(1)$ belongs to the giant component in G_1; (iii) the corresponding dependent node with probability $1 - \varphi_{C_{G_2}}(1)$ belongs to the giant component in G_2. When graphs G_1 and G_2 are two large ER random graph $G_p(N)$ with approximate Poisson degree distribution, we have $\varphi_D(z) = \varphi_{(D_{l^+}-1)}(z) = e^{E[D](z-1)}$. Thus, $\varphi_{C_{G_1}}(1) = u_A$ and $\varphi_{C_{G_2}}(1) = u_B$. The fraction S of nodes in the largest mutually connected component follows

$$S = q(1-u_A)(1-u_B) \tag{7.13}$$

where

$$\begin{cases} u_A = e^{-qE[D](u_A-1)(u_B-1)} \\ u_B = e^{-qE[D](u_A-1)(u_B-1)} \end{cases} \tag{7.14}$$

The derivative of the largest mutually connected component with respect to q in (7.13) is

$$\frac{dS}{dq} = (1-u_A)(1-u_B) - q\left[(1-u_B)\frac{du_A}{dq} + (1-u_A)\frac{du_B}{dq}\right]$$

The derivative for u_A in (7.12) follows as

$$\frac{du_A}{dq} = \frac{-\varphi'_{(D_{i^+}-1)}(u_A)(1-u_A)^2}{1-2q(1-u_A)\varphi'_{(D_{i^+}-1)}(u_A)}$$

For ER random graphs, we have that $\varphi'_{(D_{i^+}-1)}(u_A) = E[D]u_A$. Thus,

$$\frac{du_A}{dq} = \frac{-E[D]u_A(1-u_A)^2}{1-2qu_A(1-u_A)E[D]}$$

With $(1-u_A)(1-u_B) = \frac{S}{q}$ and $u_A = u_B$ from (7.14), we arrive at

$$\frac{dS}{dq} = \frac{S}{q\left(1-2E[D]\left(\sqrt{Sq}-S\right)\right)} \tag{7.15}$$

Figure 7.9 shows the straight line $y = -\frac{dS}{dq}\Big|_{1-q=\frac{1}{N}}(1-q)+1$ with slope computed from (7.15) and simulations of the largest mutually connected component for coupled ER random graphs $G_p(N)$. Again, the straight line with slope $-\frac{dS}{dq}\Big|_{1-q=\frac{1}{N}}$ shows a good estimation for the largest mutually connected component when a small fraction $1-q$ of nodes is removed.

FRACTION OF LARGEST MUTUALLY CONNECTED COMPONENT WITH ONE-TO-MANY INTERCONNECTIONS

Assume that a node is alive if at least one of its interdependent nodes is alive. Theorem 10 presents the fraction S_1 and S_2 of the largest mutually connected component for network G_1 and G_2, respectively.

Theorem 10. *Consider an interdependent network consisting of two graphs G_1 and G_2. The interconnection topology between graphs G_1 and G_2 is the random geometric graph. The fraction S_i ($i = 1,2$) of the largest mutually connected component as a function of $1-q$ removals is approximated by*

$$S_1 = q\left(1-\varphi_{C_{G_1}}(1)\right)\left(1-(1-p_{ij})^{(1-\varphi_{C_{G_2}}(1))N}\right) \tag{7.16}$$

$$S_2 = \left(1-\varphi_{C_{G_2}}(1)\right)\left(1-(1-p_{ij})^{q(1-\varphi_{C_{G_1}}(1))N}\right) \tag{7.17}$$

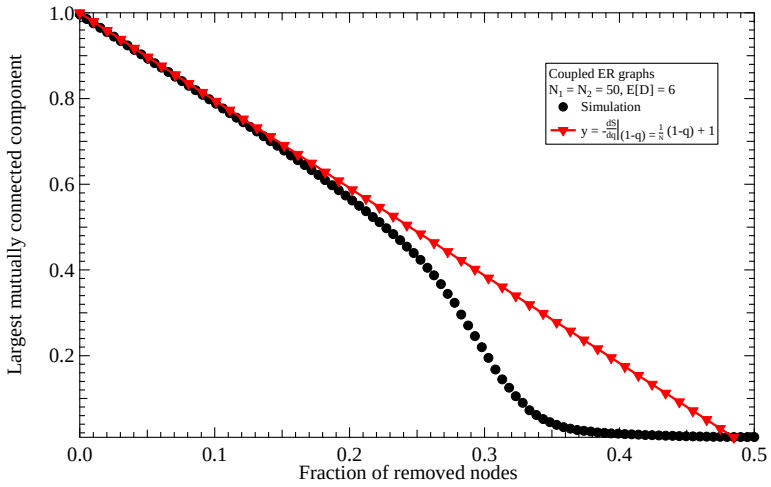


Figure 7.9: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Erdős-Rényi random graphs $G_p(N)$ with $N = 50$ and the average degree $E[D] = 6$. The interdependency is one-to-one. The results are averaged over 10^4 realizations of interdependent graphs.

with

$$\begin{cases} \varphi_{C_{G_1}}(1) = \varphi_{D_{G_1}}\left(1 - q\left(1 - (1-p_{ij})^{(1-u_B)N}\right)(1-u_A)\right) \\ \varphi_{C_{G_2}}(1) = \varphi_{D_{G_2}}\left(1 - (1 - (1-p_{ij})^{q(1-u_A)N})(1-u_B)\right) \end{cases}$$

and

$$\begin{cases} u_A = \varphi_{(D_{l^+}-1)}\left(1 - q\left(1 - (1-p_{ij})^{(1-u_B)N}\right)(1-u_A)\right) \\ u_B = \varphi_{(D_{l^+}-1)}\left(1 - (1 - (1-p_{ij})^{q(1-u_A)N})(1-u_B)\right) \end{cases}$$

where p_{ij} is the probability that there is a link l_{ij} between node i in graph G_1 and node j in graph G_2. $1-\varphi_{C_{G_1}}(1)$ is the fraction of nodes belonging to the giant component in graph G_1 and $1-\varphi_{C_{G_2}}(1)$ in graph G_2.

Proof. For network G_1, a node i is occupied with probability q. The node i is supported with at least one node with probability $1-\left(1-p_{ij}\right)^{(1-u_B)N}$ where $\left(1-p_{ij}\right)^{(1-u_B)N}$ is the probability that node i does not connect to any nodes in the giant component in graph G_2. Therefore, (7.12) is modified to

$$u_A = \varphi_{(D_{l^+}-1)}\left(1 - q\left(1 - \left(1-p_{ij}\right)^{(1-u_B)N}\right)(1-u_A)\right)$$

Analogously, for network G_2

$$u_B = \varphi_{(D_{l^+}-1)}\left(1 - (1 - (1-p_{ij})^{q(1-u_A)N})(1-u_B)\right)$$

Since we do not remove nodes from graph G_2 at the beginning of the removal, nodes in graph G_2 are occupied with probability 1. After cascading failures, a node in G_1 is in the largest mutually connected component if (i) the node is occupied with probability q; (ii) the node with probability $1-\varphi_{C_{G_1}}(1)$ belongs to the giant component in G_1; (iii) at least one of the corresponding dependent node with probability $\left(1-(1-p_{ij})^{(1-\varphi_{C_{G_2}}(1))N}\right)$ belongs to the giant component in G_2. A node in G_2 is in the largest mutually connected component if (i) the node with probability $1-\varphi_{C_{G_2}}(1)$ belongs to the giant component in G_2; (iii) at least one of the corresponding dependent node with probability $\left(1-(1-p_{ij})^{q(1-\varphi_{C_{G_1}}(1))N}\right)$ belongs to the giant component in G_1. □

When graphs G_1 and G_2 are two large ER random graphs with $\varphi_D(z)=\varphi_{(D_{i^+}-1)}(z)=e^{E[D](z-1)}$, (7.16) and (7.17) can be simplified to

$$S_1 = q(1-u_A)(1-(1-p_{ij})^{(1-u_B)N}) \tag{7.18}$$

$$S_2 = (1-u_B)(1-(1-p_{ij})^{q(1-u_A)N}) \tag{7.19}$$

7

with

$$\begin{cases} u_A = e^{E[D_1]q\left(1-(1-p_{ij})^{(1-u_B)N}\right)(u_A-1)} \\ u_B = e^{E[D_2](1-(1-p_{ij})^{q(1-u_A)N})(u_B-1)} \end{cases} \tag{7.20}$$

Figures 7.10(a) and 7.10(b) show the simulation results and S_1 and S_2 in (7.18) and (7.19) in coupled ER graphs with interconnection of random geometric graph with radius $r=0.2$. Since u_A and u_B are functions of q, computing the derivatives of u_A and u_B with respect to q in (7.20) is complicated. The derivatives of S_1 and S_2 with respect to q in (7.18) and (7.19) are even more complex. Therefore, we numerically compute the derivative $\frac{dS_i}{dq}$ $(i=1,2)$ based on (7.18) and (7.19). Figures 7.10(c,d) show the simulation results and a straight line $y=-\frac{dS_i}{dq}\Big|_{1-q=\frac{1}{N}}\left(1-q\right)+1$ $(i=1,2)$. In Figures 7.10(c,d), the straight line with slope $-\frac{dS_i}{dq}$ $(i=1,2)$ obtained from Theorem 10 shows a good approximation for the simulations for a small fraction of removals.

For the assumption that a node is alive if all its dependent nodes are alive, the results are given in the Appendix D.2.

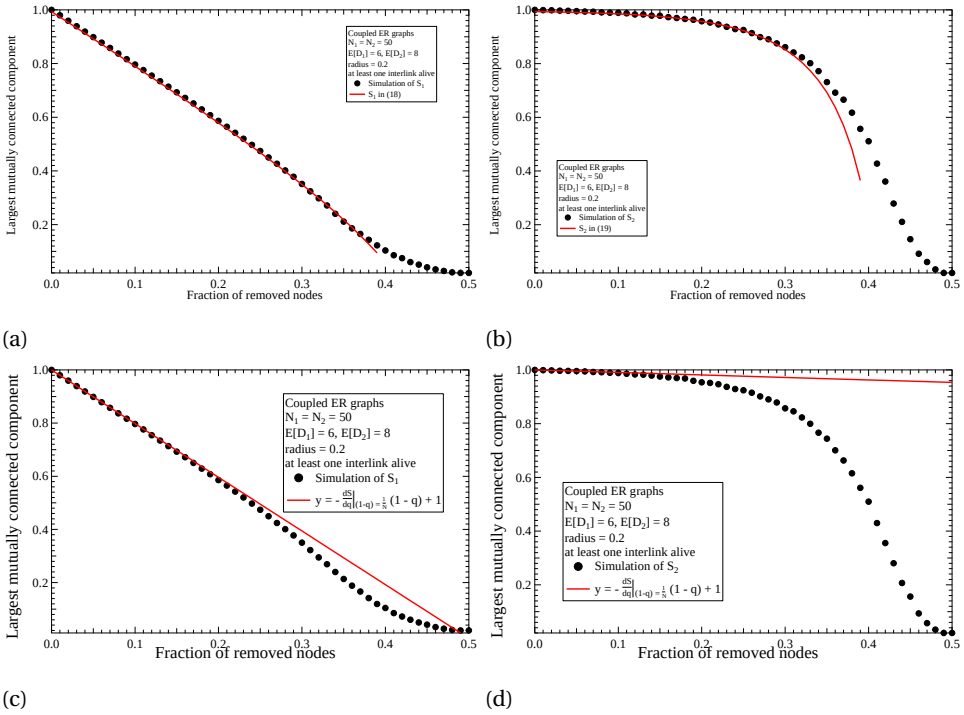


Figure 7.10: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Erdős-Rényi graphs $G_p(N)$ with $N = 50$ and the average degrees $E[D_1] = 6$ and $E[D_2] = 8$. The interconnection topology is the random geometric graph with $r = 0.2$. The results are averaged over 10^3 realizations of interdependent graphs.

7.4. SIMULATION RESULTS

In this section, we investigate the impact of two interconnection topologies, the random geometric graph and the relative neighbourhood graph, on the robustness of interdependent networks against cascading failures. The robustness is quantified by the largest mutually connected component (LMCC) when a fraction $1-q$ of nodes are removed.

We simulate a two-fold interdependent network consisting of two Erdős-Rényi (ER) graphs $G_p(N)$ or two Barabási-Albert (BA) graphs. We consider two scenarios for a node being supported by the coupled network: (i) at least one dependent nodes alive and (ii) all the dependent nodes alive. Each node has randomly assigned coordinates $0 \leq x_i \leq 1$ and $0 \leq y_i \leq 1$.

7.4.1. RANDOM GEOMETRIC GRAPH AS INTERCONNECTION

The interconnection topology between two graphs is the random geometric graph with radius r. Figure 7.11 shows the largest mutually connected component as a function of the fraction $1-q$ of the removed nodes from G_1. The interdependent network consists of two Erdős-Rényi graphs $G_p(N)$ with $N = 50$ and the average degree $E[D] = 6$. We assume a node is supported by its interconnected nodes when at least one of the interconnected nodes is alive.

For a given radius r, the LMCC in Figure 7.11 firstly decreases almost linearly with the increase of the fraction of removed nodes. Then, the LMCC experiences a first-order phase transition which is different from second-order phase transition in a single network also observed in [35] with one-to-one interconnection. Moreover, the largest mutually connected component decreases with the decrease of the radius r. For example, when a fraction 0.2 of nodes are removed, we have LMCC = 0.79 for $r = \sqrt{2}$ and LMCC = 0.69 for $r = 0.1$. The reason is that with the decrease of r, a node tends to have less interconnection nodes which increases the probability for a node to fail due to the failures of its interconnection nodes.

Figure 7.12 shows the largest mutually connected component as a function of the fraction of the removed nodes in coupled Barabási-Albert graphs. We assume a node alive when at least one of the interconnected nodes is alive. Coupled BA graphs have less distinguishable LMCC for different radius r compared to coupled ER graphs. The reason is two-fold: (i) BA graphs are robust to random failures; (ii) When we increase the radius r, a node tends to have more than one interconnections.

Figure 7.13 shows the largest mutually connected component as a function of the fraction of the removed nodes in coupled Erdős-Rényi graphs. A node is alive when all of the interconnected nodes are alive. The LMCC in Figure 7.13 decreases dramatically fast with the increase of the fraction of removed nodes. With the increase of the radius r, LMCC decreases even faster. When $r = 0.2$, the failure of 2% of the nodes collapses the whole interdependent network.

Figure 7.14 shows the largest mutually connected component as a function of the fraction of the removed nodes in coupled Barabási-Albert graphs. A node is alive when all of the interconnected nodes are alive. For a small radius r, LMCC decreases slowly with the increase of the fraction of removed nodes because (i) BA graphs are

robust to random failures; (ii) failures are less likely propagating to another network with small interconnections resulting from small r. However, for a larger radius r, LMCC decreases fast with the increase of removals $1-q$.

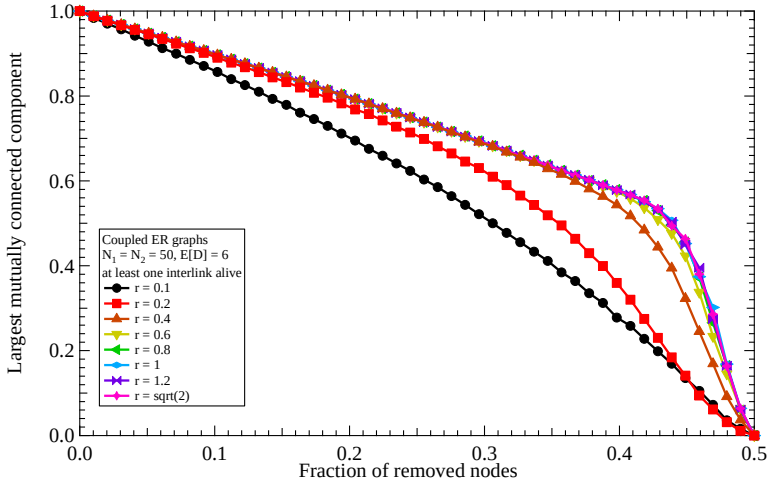


Figure 7.11: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Erdős-Rényi graphs $G_p(N)$ with $N = 50$ and the average degree $E[D] = 6$. The radius r in the random geometric graph is ranging from 0.1 to $\sqrt{2}$. The simulations are averaged over the results from 1000 interdependent graphs.

7.4.2. RELATIVE NEIGHBOURHOOD GRAPH AS INTERCONNECTION

To compare the interconnection structure of the relative neighbourhood graph and the random geometric graph, we simulate the two topologies with the same interlink density derived in Theorems 8 and 9. Figures 7.15 and 7.16 show the largest mutually connected component as a function of the fraction $1-q$ of the removed nodes in interdependent networks. The interdependent network consists of two Erdős-Rényi graphs with $N = 50$ and the average degree $E[D] = 6$ in Figure 7.15 and consists of two Barabási-Albert graphs with $N = 500$ and the average degree $E[D] = 6$ in Figure 7.16.

For both the assumptions of at least one interdependent node alive and all interdependent nodes alive, Figure 7.15 shows that the interconnection structure of the random geometric graph is more robust compared to that of the relative neighbourhood graph. An explanation is that interconnected links are evenly distributed

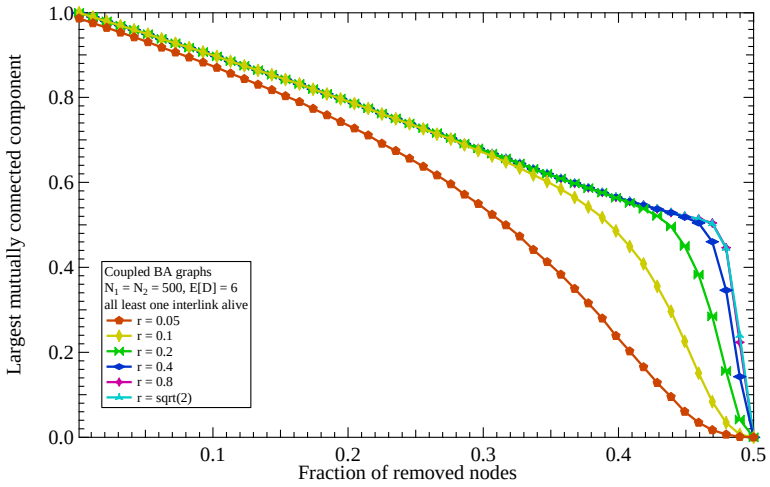


Figure 7.12: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Barabási-Albert with $N = 500$ and the average degree $E[D] = 6$. The radius r in the random geometric graph is ranging from 0.05 to $\sqrt{2}$. The results are averaged over 10^3 realizations of interdependent graphs.

in relative neighbourhood graph, whereas in random geometric graph, the interconnected links might be highly connect to few nodes depending on the location of nodes.

In Figure 7.16, the interdependent graph with coupled BA graphs shows comparable results with coupled ER graphs. Random geometric graph performs much better than relative neighbourhood graph when at least one interlinks alive. For the assumption of all interlinks alive, random geometric graph is also more robust than relative neighbourhood graph.

7.4.3. REAL-WORLD NETWORKS

To demonstrate the effectiveness of the two interconnection topologies, we interconnect two real-world coupled infrastructures in Italy [164, 176] by the random geometric graph and the relative neighbourhood graph and investigate their robustness under cascading failures.

One network is the Italian high-bandwidth backbone of the Internet consisting of $N = 39$ nodes and $L = 50$ links. The other network is the Italian high-voltage electrical transmission network consisting of $N = 310$ nodes and $L = 347$ links (ex-

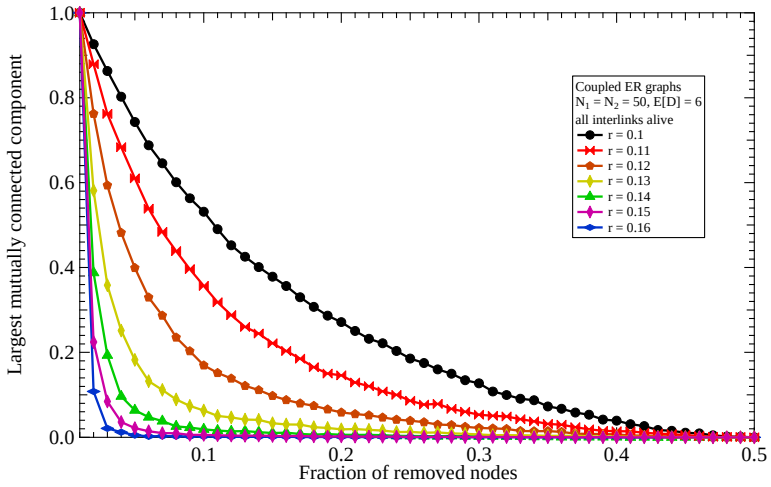


Figure 7.13: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Erdős-Rényi graphs $G_p(N)$ with $N = 50$ and the average degree $E[D] = 6$. The radius r in the random geometric graph is ranging from 0.1 to 0.16. The results are averaged over 10^3 realizations of interdependent graphs.

cluding the double links). Given the geographical locations of the nodes in the Internet and in the electrical network, we generate interconnection topologies of the random geometric graph and the relative neighbourhood graph as shown in Figures 7.17 and 7.18.

Figure 7.19 shows the largest mutually connected component as a function of the fraction of removed nodes in coupled real-world networks. The interconnection topologies are the random geometric graph and the relative neighbourhood graph with the same link density. For the assumption of at least one interlink alive, Figure 7.19 shows that the interconnection topology of the random geometric graph is more robust than that of the relative neighbourhood graph. However, the relative neighbourhood graph is more robust than the random geometric graph for the assumption of all interlinks alive.

7.5. CONCLUSION

In this paper, we investigate two interconnection topologies for interdependent networks that incorporate the locations of nodes. The two topologies generalize the

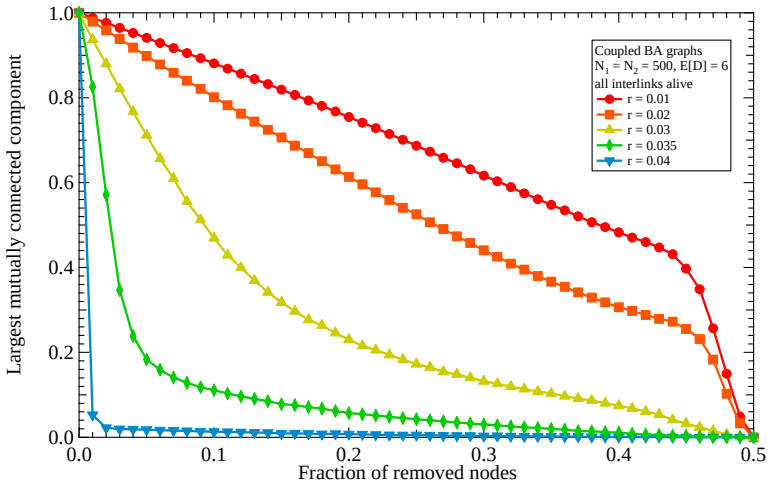


Figure 7.14: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Barabási-Albert with $N = 500$ and the average degree $E[D] = 6$. The radius r in the random geometric graph is ranging from 0.01 to 0.04. The results are averaged over 10^3 realizations of interdependent graphs.

one-to-one interconnection to an arbitrary number of interconnections depending on the locations of nodes. We analyse the properties of the two topologies and the impact of the two interconnection structures on robustness of interdependent networks against cascading failures. Specifically, the derivation of the number of links in the two topologies enables the comparison of robustness performance between the two topologies. In particular, the random geometric graph provides the flexibility for network providers to determine the link density of interconnection in order to achieve the desired robustness level. The relative neighbourhood graph, often used in wireless networks [177] to provide optimal coverage with least energy consumption, as an interconnection structure is less robust compared to the random geometric graph.

In addition, we propose the derivative of the largest mutually connected component as a new robust metric which addresses the impact of a small fraction of failed nodes. To avoid the collapse of the whole network, the proposed robustness metric quantifies the damage of networks triggered by a small fraction of failures, significantly smaller than the fraction at the critical threshold, that corresponds to the collapse of the whole network.

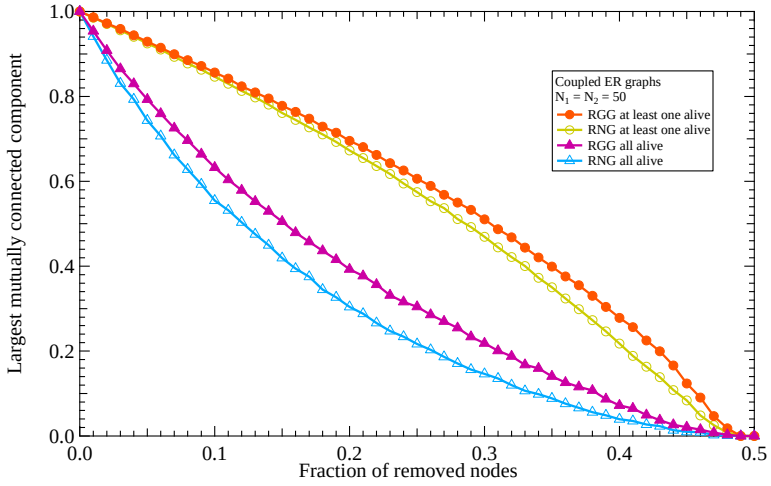


Figure 7.15: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Erdős-Rényi graphs $G_p(N)$ with $N = 50$ and the average degree $E[D] = 6$. The interconnection topology is the relative neighbourhood graph. The results are averaged over 10^3 realizations of interdependent graphs.

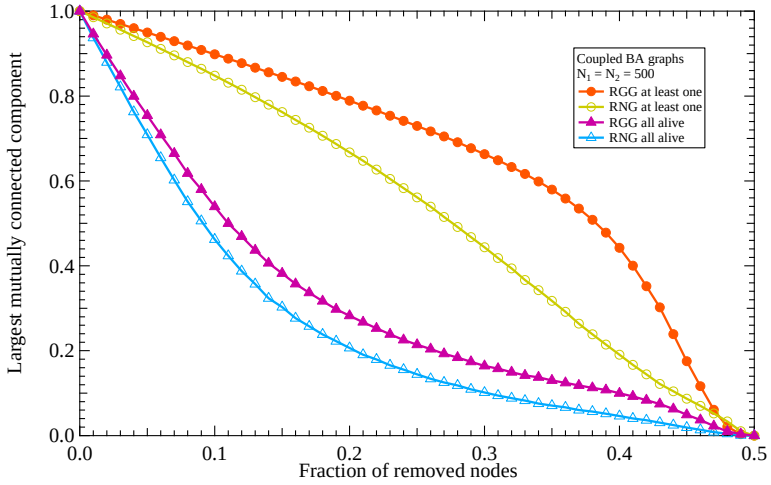


Figure 7.16: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Barabási-Albert graphs with $N = 500$ and the average degree $E[D] = 6$. The interconnection topology is the relative neighbourhood graph. The results are averaged over 10^3 realizations of interdependent graphs.

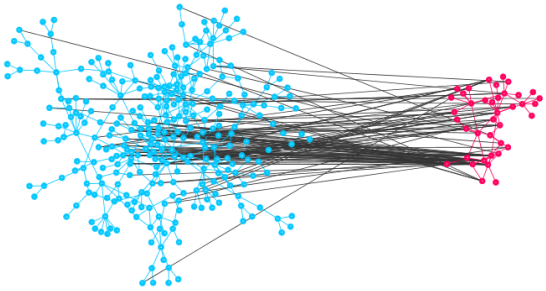


Figure 7.17: Coupled Italian electrical transmission network (blue) and the Italian backbone of the Internet (red) with the interconnection topology of the random geometric graph.

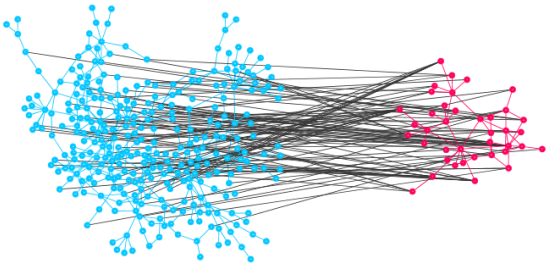


Figure 7.18: Coupled Italian electrical transmission network (blue) and the Italian backbone of the Internet (red) with the interconnection topology of the relative neighbourhood graph.

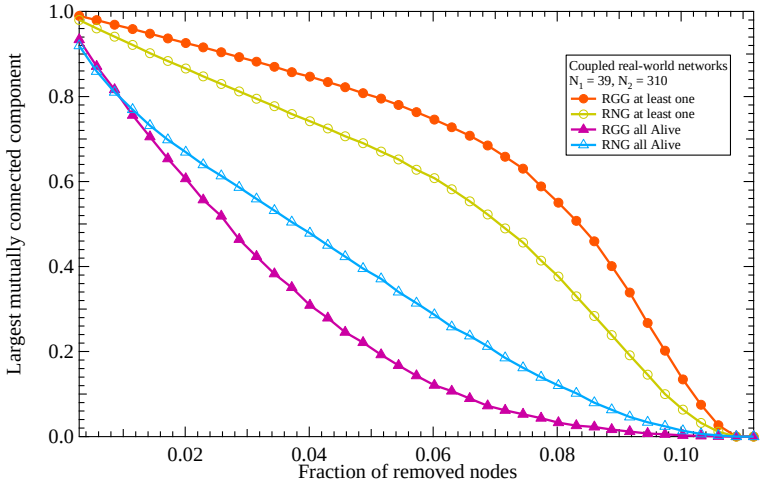


Figure 7.19: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are the Italian high-bandwidth backbone of the Internet and the Italian high-voltage electrical transmission network. The interconnection topologies are the random geometric graph and the relative neighbourhood graph with the same link density.

8

STRUCTURAL TRANSITION IN INTERDEPENDENT NETWORKS WITH REGULAR INTERCONNECTIONS

8.1. INTRODUCTION

An interdependent network, also called an interconnected network or a network of networks, is a network consisting of different types of networks that depend upon each other for their functioning [178]. For example, power networks dependent on communication networks, where each node in a communication network controls one or more nodes in a power network, while each communication node needs power to function [164]. Critical infrastructures, such as telecommunications, power systems, transportation, water/oil/gas-supply systems, are highly interconnected and mutually depend upon each other. Due to the interdependencies between infrastructures, Little [179] proposed to view infrastructures as systems of systems to understand their robustness against cascading failures. Disasters like

large-scale blackouts have shown that most vulnerability lies in the interdependencies between different infrastructures which allow the failure in one infrastructure to propagate to another infrastructure [180].

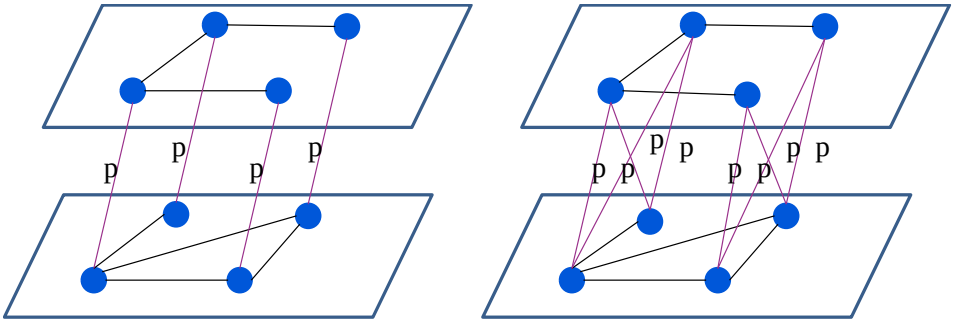
The coupling between networks can modify the dynamical processes running on interdependent networks. For example, Buldyrev *et al.* [35] show that the collapse of interdependent networks occurs abruptly while the collapse of individual networks is approached continuously. The epidemic threshold for epidemic spreading processes is characterized by both the topologies of each coupled network and the interconnection topology between them [36, 181].

Radicchi and Arenas [182] motivated the use of an interdependent model consisting of two connected networks, G_1 and G_2, with weighted interconnection links. The coupling weight between two networks is determined by a non-negative real value p. In coupled electrical and communication networks, the weight can be interpreted as the power dispatched by the electrical node. Radicchi and Arenas [182] and also Martin-Hernandez *et al.* [183] found the existence of a structural transition point p^* that separates an interdependent network into two regimes: for $p > p^*$, the interdependent network acts as whole, whereas for $p < p^*$, the network is structurally separated as graphs G_1 and G_2. The explicit expression for the transition threshold p^* is determined in [184].

However, the model of Radicchi and Arenas [182] is limited to a one-to-one interconnection which means that one node in graph G_1 connects to one and only one node in graph G_2 and vice versa. When the interconnection pattern is not one-to-one, i.e. $B \neq pI$, as in most real-world examples, the determination of the transition threshold p^* is more complex. Examples for a multiple-to-multiple interconnection pattern rather than a one-to-one interconnection can be found in (i) smart grids consisting of coupled sensor networks and power networks [15, 185, 186] where a sensor might control multiple power stations due to cost and energy budget; (ii) functional brain networks modelled as multi-layer network where one brain region in one layer can exert influence over any node in the other layer [187]; (iii) infrastructures like power systems and fiber-optic communication systems that are geographically interconnected based on spatial proximity [188].

In this paper, we investigate the structural threshold p^* in interdependent networks with a general k-to-k (k is a positive integer) interconnection as shown in Fig-

ure 8.1. We derive an upper bound for the structural threshold p^*. We find certain topologies where the upper bound is reached. We interpret the physical meaning of the structural threshold p^* with respect to the minimum cut. For some special cases, an analytical expression for the structural threshold p^* is presented. Furthermore, we show with a counter example that the structural threshold p^* does not always exist.



(a) one-to-one interconnection (b) a general k-to-k interconnection with $k = 2$

Figure 8.1: We generalize the model of Radicchi and Arenas from a one-to-one interconnection to a general k-to-k interconnection.

The paper is organized as follows. Section 8.2 introduces interdependent networks. Section 8.3 provides an upper bound for the structural threshold p^* for a general k-to-k interconnection pattern. Section 8.4 interprets the physical meaning of the structural threshold p^*. Section 8.5 derives the exact structural threshold p^* for special cases of the interconnection and presents a counter example for the non-existence of the structural threshold p^*. Section 8.6 concludes the paper.

8.2. AN INTERDEPENDENT NETWORK

Let the graph $G(N, L)$ represent an interdependent network consisting of two networks, represented by graph G_1 with n nodes and graph G_2 with m nodes. The total number of nodes in G is $N = n + m$. An interdependency link connects a node i in network G_1 to a node j in network G_2. The adjacency matrix A of the interdependent network G has the block structure

$$A = \begin{bmatrix} (A_1)_{n\times n} & B_{n\times m} \\ \left(B^T\right)_{m\times n} & (A_2)_{m\times m} \end{bmatrix}$$

where A_1 is the $n \times n$ adjacency matrix of G_1, A_2 is the $m \times m$ adjacency matrix of G_2 and B is the $n \times m$ interconnection matrix representing the interconnections between G_1 and G_2. If each interdependent link is weighted with a non-negative real number p, the matrix B is a weighted matrix with elements $b_{ij} = p$ if node i in G_1 connects to node j in G_2, otherwise $b_{ij} = 0$. The definition for B in [178] is more general where the weight on each interdependent link can be different. Here, matrix B corresponds to the scenario that each interdependent link has a weight of p.

A k-**to-**k **interconnection**, where $k = 1, 2, \cdots, \min(n, m)$, means that one node in graph G_1 connects to k nodes in graph G_2 and vice-versa. We only consider undirected interconnection links. The k-to-k interconnection requires a square interconnection matrix B with $n = m$, because the number kn of interconnection links computed in graph G_1 must equal to the number km computed in graph G_2, i.e., $kn = km$. In the rest of this article, we focus on a square interconnection matrix B with $n = m$ and the subscript of matrix B is omitted. The k-to-k interconnection is a generalization of the one-to-one interconnection ($B = pI$) studied in [35, 182, 184].

For a square interconnection matrix B, a k-to-k interconnection can be constructed via a circulant matrix [42] with the form

$$B = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & \cdots & n \end{array} \\ \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_n \\ c_n & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_n & c_1 & \cdots & c_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ c_2 & c_3 & c_4 & \cdots & c_1 \end{bmatrix} \end{array} \tag{8.1}$$

where the row vector $(c_1, c_2, \ldots, c_n)$ has exactly k elements of p and $n-k$ elements of 0. Each row and each column of B contains the same number of non-zero elements, but the position of the non-zero elements is shifted. For example, a symmetric ma-

trix B for a 2-to-2 ($k = 2$) interconnection can be written as

$$B = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & \cdots & n \end{array} \\ \begin{bmatrix} 0 & p & 0 & \cdots & p \\ p & 0 & p & \cdots & 0 \\ 0 & p & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ p & 0 & 0 & \cdots & 0 \end{bmatrix} \end{array}$$

Analogous to the definition of the Laplacian matrix $Q = \Delta - A$ in a single network, where Δ is the diagonal matrix of node degrees, we use the following diagonal matrices:

$$\Delta_1 \overset{def}{=} \text{diag}(Bu)$$

$$\Delta_2 \overset{def}{=} \text{diag}\left(B^T u\right)$$

to define the **Laplacian matrix Q of the interdependent network** G as

$$Q = \begin{bmatrix} Q_1 + \Delta_1 & -B \\ -B^T & Q_2 + \Delta_2 \end{bmatrix}$$

where Q_1 and Q_2 are the Laplacian matrices of networks G_1 and G_2, respectively. The all-one vector is denoted by u and the subscript of u is used if the dimension is not clear. Since the Laplacian matrix Q is symmetric, the eigenvalues of Q are non-negative and at least one is zero [42]. We order the eigenvalues of the Laplacian matrix Q as $0 = \mu_N \leq \mu_{N-1} \leq \cdots \leq \mu_1$ and denote the eigenvector corresponding to the k-largest eigenvalue by x_k. The second smallest eigenvalue of the Laplacian matrix Q is coined by Fiedler [189] as the algebraic connectivity μ_{N-1} of a graph G. The algebraic connectivity plays a key role in different aspects of the robustness of networks, such as diffusion processes [183, 190], synchronization stability [191] and network robustness against failures [156].

The Laplacian eigenvalue equation for the eigenvector $x_k = \left(x_1^T, x_2^T\right)^T$, where x_1 and x_2 are $n \times 1$ vectors, belonging to the eigenvalue μ_k is

$$\begin{bmatrix} Q_1 + \Delta_1 & -B \\ -B^T & Q_2 + \Delta_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mu_k \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8.2}$$

The normalized vector $x_N = \frac{1}{\sqrt{N}}\left(u_n^T,\ u_n^T\right)^T$ is an eigenvector belonging to the smallest eigenvalue $\mu_N = 0$ of the Laplacian Q. We briefly present a theorem in [178, Theorem 3] to introduce a non-trivial eigenvalue and eigenvector of the Laplacian Q.

Theorem 11. *Only if the $n \times m$ interconnection matrix $\widetilde{B}$ has a constant row sum equal to $\frac{\mu^*}{N}m$ and a constant column sum equal to $\frac{\mu^*}{N}n$, which we call the regularity condition for $\widetilde{B}_{n\times m}$,*

$$\begin{cases} \widetilde{B}u_m = \frac{\mu^*}{N}mu_n \\ \widetilde{B}^T u_n = \frac{\mu^*}{N}nu_m \end{cases}$$

then is

$$x = \frac{1}{\sqrt{N}}\left[\sqrt{\frac{m}{n}}u_n^T,\ -\sqrt{\frac{n}{m}}u_m^T\right]^T$$

an eigenvector of Q belonging to the eigenvalue

$$\mu^* = \left(\frac{1}{n}+\frac{1}{m}\right)u_n^T\widetilde{B}_{n\times m}u_m$$

and $u_n^T\widetilde{B}_{n\times m}u_m$ equals the sum of the elements in $\widetilde{B}$, representing the total strength of the interconnection between graphs G_1 and G_2.

Corollary 2. *Consider an interdependent graph G with N nodes consisting of two graphs each with n nodes, whose interconnections are described by a weighted interconnection matrix B. For a k-to-k interconnection pattern with the coupling weight p on each interconnection link, the vector*

$$x = \frac{1}{\sqrt{N}}\left[u_n^T,\ -u_n^T\right]^T \tag{8.3}$$

is an eigenvector of the Laplacian matrix Q of graph G belonging to the eigenvalue

$$\mu^* = 2kp \tag{8.4}$$

Proof. For a k-to-k interconnection, the row and column sum of the interconnection matrix B is a constant which equals to kp,

$$\begin{cases} Bu_n = kpu_n \\ B^T u_n = kpu_n \end{cases}$$

which obeys the regularity condition in Theorem 11. With $n = m$ and the total coupling strength $u_n^T\widetilde{B}_{n\times m}u_m = kpn$ in Theorem 11, we establish the Corollary 2. □

8

The coupling weight p on each interconnection link can be varied from 0 to ∞. Corollary 11 implies that there is a value of $p > 0$ for which $\mu^* = 2kp$ in (8.4) can be made the smallest positive eigenvalue, which then equals the algebraic connectivity μ_{N-1} of the whole interdependent network G. By increasing the coupling weight p, the non-trivial eigenvalue $\mu^* = 2kp$ is no longer the second smallest eigenvalue. There exists a transition threshold p^* such that $\mu_{N-1} \neq 2kp$ when $p > p^*$. Because the eigenvalues of Laplacian Q are continuous functions of the coupling weight p, the second and third smallest eigenvalue coincide [184] at the point of the transition threshold p^*.

The Laplacian matrix Q for a k-to-k interconnection can be written as the sum of two matrices $Q = \begin{bmatrix} Q_1 & O \\ O & Q_2 \end{bmatrix} + \begin{bmatrix} kpI & -B \\ -B^T & kpI \end{bmatrix}$. According to the interlacing theorem for the sum of two matrices [42], a lower bound for the third smallest eigenvalue μ_{N-2} of the Laplacian matrix Q follows

$$\mu_{N-2}(Q) \geq \min(\mu_{n-2}(Q_1),\ \mu_{n-2}(Q_2)) \tag{8.5}$$

where $\mu_{n-2}(Q_1)$ and $\mu_{n-2}(Q_2)$ are the third smallest eigenvalue of graphs G_1 and G_2, respectively.

8.3. AN UPPER BOUND FOR THE TRANSITION THRESHOLD p^*

This section derives an upper bound for the transition threshold p^* of interdependent networks with k-to-k interconnection patterns. We find topologies for interdependent networks, where the upper bound is attained.

8.3.1. UPPER BOUND FOR p^*

For a given interconnection matrix B with a k-to-k interconnection, i.e. $Bu = B^T u = kpu$, the Laplacian matrix Q is written as

$$Q = \begin{bmatrix} Q_1 + kpI & -B \\ -B^T & Q_2 + kpI \end{bmatrix} \tag{8.6}$$

For any normalized vector $x = \left(x_1^T,\ x_2^T\right)^T$, the quadratic form $x^T Q x$ of the Laplacian Q follows

$$x^T Q x = kp + x_1^T Q_1 x_1 + x_2^T Q_2 x_2 - 2x_1^T B x_2 \tag{8.7}$$

Let x_1 be an eigenvector corresponding to the second smallest eigenvalue $\mu_{n-1}(Q_1)$ of Q_1 and $x_2 = 0$. For vector $x = \left(x_1^T,\ 0\right)^T$, the normalization of vector x reads $x^T x = x_1^T x_1 = 1$. Thus, the quadratic form in (8.7) follows $x^T Q x = kp + \mu_{n-1}(Q_1)$. Analogously, we have $x^T Q x = kp + \mu_{n-1}(Q_2)$ when $x_1 = 0$ and x_2 be the eigenvector belonging to $\mu_{n-1}(Q_2)$. Applying the Rayleigh inequality [42] to the algebraic connectivity μ_{N-1} yields

$$\mu_{N-1} \leq \frac{x^T Q x}{x^T x}$$

With vector $x = \left(x_1^T,\ 0\right)^T$ or vector $x = \left(0,\ x_2^T\right)^T$, we arrive at

$$\mu_{N-1} \leq \min\left(\mu_{n-1}(Q_1),\ \mu_{n-1}(Q_2)\right) + kp \tag{8.8}$$

Equality holds when x is the eigenvector belonging to the algebraic connectivity μ_{N-1}.

The non-trivial eigenvalue $\mu^* = 2kp$ in (8.4) corresponding to eigenvector $x = \frac{1}{\sqrt{N}}\left(u_n^T,\ -u_n^T\right)^T$ can be made the algebraic connectivity μ_{N-1} for $p < p^*$, whereas μ^* is no longer the algebraic connectivity μ_{N-1} for $p > p^*$. At the transition threshold p^*, the algebraic connectivity is $\mu_{N-1} = 2kp^*$. Substituting $\mu_{N-1} = 2kp^*$ and $p = p^*$ in (8.8), we arrive at an upper bound for the transition threshold p^*

$$p^* \leq \frac{1}{k}\min(\mu_{n-1}(Q_1),\ \mu_{n-1}(Q_2)) \tag{8.9}$$

Figures 8.2 and 8.3 show the accuracy of the upper bound (8.9) in interdependent networks with size $N = 1000$ consisting of two Erdős-Rényi graphs $G_q(n)$ with link density q and two Barabási-Albert graphs with average degree $d_{av} = 6$. The interconnection pattern is 2-to-2 ($k = 2$). The upper bound in Figure 8.2 provides a good approximation for the transition threshold p^*. The upper bound in Figure 8.3 is less accurate than that in Figure 8.2.

Radicchi and Arenas [182] show that the transition threshold p^* is upper bounded by $p^* \leq \frac{1}{4}\mu_{N-1}(Q_1 + Q_2)$ when $B = pI$ (the k-to-k interconnection with $k = 1$). The exact p^* is determined in [184] when $B = pI$. However, the method in [184] cannot be readily generalized to a two-to-two nor to a general k-to-k ($k \neq 1$) interconnection.

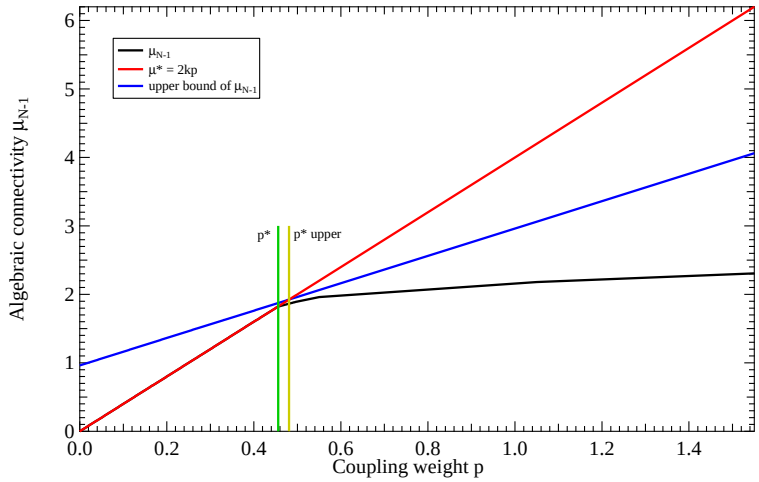


Figure 8.2: Accuracy of the bound for the transition threshold p^* in interdependent networks consisting of two Erdős-Rényi graphs $G_q(n)$ with $n = 500$ and average degree $d_{av} = 6$. The interconnection pattern is 2-to-2, i.e., $k = 2$.

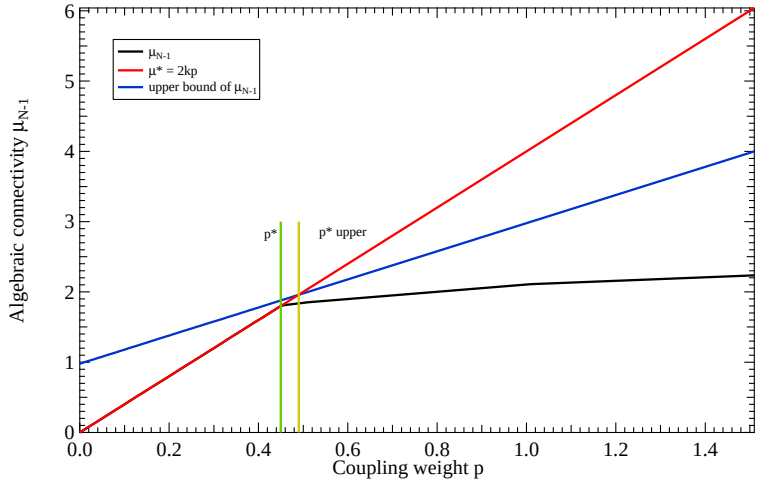


Figure 8.3: Accuracy of the bound for the transition threshold p^* in interdependent networks consisting of two Barabási-Albert graphs with $n = 500$ and average degree $d_{av} = 6$. The interconnection pattern is 2-to-2, i.e., $k = 2$.

8.3.2. TOPOLOGIES FOR WHICH THE UPPER BOUND (8.9) IS EXACT

An interesting question is when vectors $x = \left(x_1^T,\ 0\right)^T$ and $x = \left(0,\ x_2^T\right)^T$ are eigenvectors of Q belonging to eigenvalues of $\mu_{n-1}(Q_1) + kp$ and $\mu_{n-1}(Q_2) + kp$. Theorem 12 presents two conditions for vector x to be an eigenvector of the Laplacian Q. We

firstly introduce two definitions. A graph is defined as a singular graph if its adjacency matrix has at least one zero eigenvalue. A kernel eigenvector is an eigenvector of a singular graph belonging to the zero eigenvalue [192].

Theorem 12. *Vector* $x = \left(x_1^T,\, 0\right)^T$ *is an eigenvector of* Q *belonging to the eigenvalue* $\mu_{n-1}(Q_1) + kp$ *if*

(i) x_1 *is the eigenvector corresponding to the second smallest eigenvalue* $\mu_{n-1}(Q_1)$ *of* Q_1*;*

and

(ii) x_1 *is the kernel eigenvector of matrix* B^T*, i.e.,* $B^T x_1 = 0$.

Proof. Assuming a k-to-k interconnection pattern, the Laplacian eigenvalue equation (8.2) for $x = \left(x_1^T,\, 0\right)^T$ reads

$$\begin{bmatrix} Q_1 + pkI & -B \\ -B^T & Q_2 + pkI \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 x_1 + kpx_1 \\ -pB^T x_1 \end{bmatrix}$$

Conditions (i) and (ii) yield $Q_1 x_1 = \mu_{n-1}(Q_1) x_1$ and $B^T x_1 = 0$. Thus, we have

$$\begin{bmatrix} Q_1 + pkI & -pB \\ -pB^T & Q_2 + pkI \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = \left(\mu_{n-1}(Q_1) + kp\right) \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

from which we establish Theorem 12. □

8

Analogously, Theorem 12 identifies vector $x = \left(0,\, x_2^T\right)^T$ as an eigenvector belonging to $\mu_{n-1}(Q_2) + pk$ if x_2 satisfies conditions (i) and (ii).

We now present the topology of an interdependent graph G consisting of graphs G_1 and G_2, where Theorem 12 holds and the upper bound (8.9) for transition threshold p^* is attained. Without loss of generality, we assume that graph G_1 has a smaller algebraic connectivity $\mu_{n-1}(Q_1)$ than $\mu_{n-1}(Q_2)$. We construct a graph G_1 and an interconnection matrix B where conditions (i) and (ii) in Theorem 12 are satisfied. Graph G_2 can be any topology with $\mu_{n-1}(Q_2) \geq \mu_{n-1}(Q_1)$.

The join [193] of two graphs H_1 and H_2 with adjacency matrices A_{H_1} and A_{H_2}, denoted as $H_1 \vee H_2$, is a graph consisting of graphs H_1 and H_2 where each node in H_1 is connected to each node in H_2 as shown in Figure 8.4. The adjacency matrix of the join graph $H_1 \vee H_2$ has the block form $\begin{bmatrix} \left(A_{H_1}\right)_{n\times n} & J_{n\times m} \\ \left(J^T\right)_{m\times n} & \left(A_{H_2}\right)_{m\times m} \end{bmatrix}$, where $J_{n\times m}$ is the

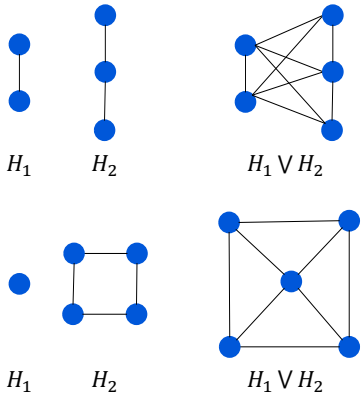


Figure 8.4: Two examples for the join of two graphs

$n \times m$ all-one matrix. The join operation on two graphs is useful in determining the synchronizability of complex networks [194].

For a k-to-k interconnection pattern, we divide graph G_1 on n nodes into $\frac{n}{k}$ subgraphs $H_1, \ldots, H_{\frac{n}{k}}$ where each subgraph has exactly k nodes and $\frac{n}{k}$ is an integer. In other words, k is chosen in such a way that $k \mid n$, i.e., k is a divisor of n. The $k \times k$ adjacency matrix for a subgraph H_i is denoted by A_{H_i}. Graph G_1 is constructed as the join of $\frac{n}{k}$ graphs $G_1 = H_1 \vee H_2 \vee \ldots \vee H_{\frac{n}{k}}$. The adjacency matrix A_1 of graph G_1 can be written as a block matrix

$$A_1 = \begin{bmatrix} A_{H_1} & J & & \\ J & A_{H_2} & J & \\ & \ddots & \ddots & \ddots \\ & & J & A_{H_{\frac{n}{k}}} \end{bmatrix} \tag{8.10}$$

To obey the condition (i) in Theorem 12, we focus on the eigenvector x_1 belonging to the algebraic connectivity of graph G_1. A theorem proved in [195], is introduced for the Laplacian eigenvalues of the join of two graphs.

Theorem 13. *Let G_1 and G_2 be graphs on n and m nodes, respectively. If $\mu_1, \mu_2, \ldots, \mu_n$ are the Laplacian eigenvalues of graph G_1 and $\alpha_1, \alpha_2, \ldots, \alpha_m$ are the Laplacian eigenvalues of graph G_2, then the Laplacian eigenvalues of the join $G_1 \vee G_2$ are $m+n, \mu_1+m, \ldots, \mu_{n-1}+m, \alpha_1+n, \ldots, \alpha_{m-1}+n$ and 0. Suppose that y is an eigenvector of G_1*

that is orthogonal to the all-one vector u_n. Extend y to a vector of size $m+n$ by defining the m components to be zero. If y is an eigenvector belonging to the eigenvalue μ, the extension of y is an eigenvector of $G_1 \vee G_2$ belonging to the eigenvalues $\mu + m$. Similarly, an eigenvector of the eigenvalue α in G_2 extends to an eigenvector of $G_1 \vee G_2$ belonging to the eigenvalue $\alpha+n$. The eigenvalue $m+n$ corresponds to the eigenvector $x = \frac{1}{\sqrt{m+n}} \left[\sqrt{\frac{m}{n}} u_n^T, -\sqrt{\frac{n}{m}} u_m^T \right]^T$.

Theorem 13 can be generalized from the join of two graphs to the join of $\frac{n}{k}$ graphs. The Laplacian eigenvalues of graph $G_1 = H_1 \vee H_2 \vee \ldots \vee H_{\frac{n}{k}}$ are 0, n and the Laplacian eigenvalues $\mu(H_i)$ of each subgraph H_i plus k. The eigenvector for eigenvalue $\mu(H_i) + k$ is the extension of eigenvector $x(H_i)$ belonging to the eigenvalue $\mu(H_i)$ in subgraph H_i.

The algebraic connectivity μ_{n-1} of graph G_1 equals

$$\mu_{n-1} = \min\left(\mu_{k-1}(H_1),\, \mu_{k-1}(H_2),\, \ldots,\, \mu_{k-1}\left(H_{\frac{n}{k}}\right)\right) + k$$

where $\mu_{k-1}(H_i)$ is the second smallest eigenvalue of subgraph H_i on k nodes. Particularly, if we assume that subgraph H_1 has the smallest algebraic connectivity after node relabelling, then the algebraic connectivity of graph G_1 is $\mu_{n-1} = \mu_{k-1}(H_1) + k$ and the corresponding eigenvector is

$$x_1 = [\left(x_{k-1}^T\right)_{1\times k},\, O_{1\times(n-k)}]^T \tag{8.11}$$

8

where x_{k-1} is the eigenvector corresponding to the second smallest eigenvalue $\mu_{k-1}(H_1)$ of subgraph H_1. The eigenvector in (8.11) is orthogonal to the all-one vector $u_{n\times 1}$ and determines the topology of the interconnection matrix B.

Next, we construct a matrix B that satisfies condition (ii) in Theorem 12, i.e., $B^T x_1 = 0$,

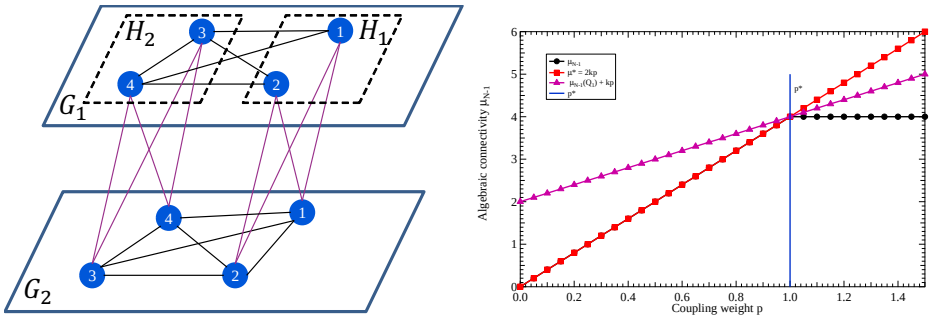
$$B^T = \begin{bmatrix} pJ_{k\times k} & O_{k\times(n-k)} \\ O_{(n-k)\times k} & C_{(n-k)\times(n-k)} \end{bmatrix} \tag{8.12}$$

where the matrix C can be a general regular matrix with $Cu_{n-k} = C^T u_{n-k} = kpu_{n-k}$. The matrix $J_{k\times k}$ in the block matrix B^T means that the nodes labelled 1, …, k in G_2 are fully connected to nodes labelled 1, …, k in G_1. The matrix $C_{(n-k)\times(n-k)}$ in the block matrix B^T means that the remaining nodes labelled $k+1$, …, n in G_2 can connect to any k nodes labelled $k+1$, …, n in G_1, and vice-versa.

Matrices A_1 in (8.10) and B in (8.12) satisfy Theorem 12, because the vector $x = \left(x_1^T, 0\right)^T$ is an eigenvector belonging to the eigenvalue $\mu_{n-1}(Q_1)+kp$ of the Laplacian Q. The lower bound in (8.5) shows that the third smallest eigenvalue $\mu_{N-2}(Q) \geq \mu_{n-2}(Q_1) \geq \mu_{n-1}(Q_1)$. The eigenvalue $\mu_{n-1}(Q_1)+kp$ can be made the third smallest eigenvalue $\mu_{N-2}(Q)$ of the Laplacian Q if the coupling weight p is small. At the transition threshold p^*, the third smallest eigenvalue $\mu_{N-2}(Q) = \mu_{n-1}(Q_1)+kp$ equals to the second smallest eigenvalue $\mu^* = 2kp$ of the Laplacian matrix Q. The exact transition threshold p^* thus follows from $\mu_{n-1}(Q_1)+kp^* = 2kp^*$ as

$$p^* = \frac{1}{k}\mu_{n-1}(Q_1) \tag{8.13}$$

Figure 8.5 shows an example with graph G_1 and interconnection matrix B constructed from (8.10) and (8.12). In Figure 8.5(a), graph G_1 is the join of graphs H_1 and H_2. The interconnection matrix with a 2-to-2 ($k = 2$) interconnection is $B = \begin{bmatrix} pJ_{2\times 2} & O \\ O & pJ_{2\times 2} \end{bmatrix}$. Figure 8.5(b) shows that the transition occurs at the point $p^* = \frac{1}{k}\mu_{n-1}(Q_1)$, as predicted in (8.13).



(a) An example graph

(b) Transition threshold p^*

Figure 8.5: An example topology where the upper bound (8.9) is exact.

8.4. PHYSICAL MEANING OF p^* IN TERMS OF THE MINIMUM CUT

In graph theory, a cut [42] is defined as the partition of a graph into two disjoint subgraphs $\widetilde{G_1}$ and $\widetilde{G_2}$. A cut set refers to a set of links between subgraphs $\widetilde{G_1}$ and $\widetilde{G_2}$. For a weighted graph, the minimum cut refers to a cut set whose cut weight R is

minimized, where the cut weight R is the sum of link weights over all links in the cut set. In this paper, the interdependent network G is weighted, where each link within graphs G_1 and G_2 has weight 1 and each link between graphs G_1 and G_2 has weight p.

A normalized index vector y for a cut of a graph G into subgraphs $\widetilde{G_1}$ and $\widetilde{G_2}$ is defined as

$$y_i = \sqrt{\frac{1}{N}} \begin{cases} 1 & \text{if node } i \in \widetilde{G_1} \\ -1 & \text{if node } i \in \widetilde{G_2} \end{cases}$$

where $y^T y = 1$. The cut weight R follows [42] from the quadratic form of the Laplacian matrix Q

$$R = \frac{Np}{4} \sum_{l \in \mathscr{L}} (y_{l^+} - y_{l^-})^2 = \frac{N}{4} y^T Q y$$

because $y_{l^+} - y_{l^-} = \frac{2}{\sqrt{N}}$ if the starting node l^+ and the ending node l^- of a link l belong to different subgraphs, otherwise $y_{l^+} - y_{l^-} = 0$. The minimum cut is [42]

$$R_{\min} = \frac{N}{4} \min_{y \in \mathbb{Y}} y^T Q y$$

where $\mathbb{Y}$ is the set of all possible normalized index vectors of the N-dimensional space. Rayleigh's theorem [42] states that, for any normalized vector y orthogonal to the all-one vector u, we have that $\mu_{N-1} \le \frac{y^T Q y}{y^T y} \le y^T Q y$ because $y^T y = 1$, and the equality holds when y is an eigenvector belonging to μ_{N-1}. With $\mu_{N-1} \le y^T Q y$, the minimum cut $R_{\min}$ follows

$$R_{\min} \ge \frac{N}{4} \mu_{N-1}$$

If the index vector y is an eigenvector of G belonging to the eigenvalue μ_{N-1}, then we obtain that $R_{\min} = \frac{N\mu_{N-1}}{4}$. Corollary 2 implies that the eigenvalue $\mu^* = 2kp$ can be made the second smallest eigenvalue μ_{N-1} with eigenvector $x = \frac{1}{\sqrt{N}} \left[u_n^T, -u_n^T\right]^T$ if $p < p^*$. If $p < p^*$, the partition corresponding to $y = x$ results in the minimum cut with $R_{\min} = \frac{N\mu_{N-1}}{4}$. The resulting subgraphs from that partition are exactly graphs G_1 and G_2 and the cut set contains all the interdependent links. When the coupling weight $p > p^*$, the eigenvector $x = \frac{1}{\sqrt{N}} \left[u_n^T, -u_n^T\right]^T$ is no longer an eigenvector of graph G belonging to the second smallest eigenvalue μ_{N-1}. The minimum cut cannot be achieved by only cutting all the interconnection links.

The physical meaning of p^* in terms of the minimum cut is that if $p < p^*$, the minimum cut can be achieved by cutting all the interconnection links, while above

p^*, the minimum cut involves both links within each subgraph and the interdependent links between two subgraphs of an interdependent network G.

8.5. EXACT THRESHOLD FOR SPECIAL STRUCTURES OF INTERDEPENDENT NETWORKS

In this section, we analytically determine the structural threshold p^* for special graphs G_1 and G_2 or a special interconnection matrix B.

8.5.1. COUPLED IDENTICAL CIRCULANT GRAPHS

Let x_{n-1} be the eigenvector belonging to the second smallest eigenvalue $\mu_{n-1}(Q_1)$ of the Laplacian matrix Q_1 of graph G_1. For vector $x = \left(x_{n-1}^T,\ x_{n-1}^T\right)^T$ and $Q_2 = Q_1$, the eigenvalue equation in (8.2) reads

$$\begin{bmatrix} Q_1 + kpI & -p\hat{B} \\ -p\hat{B}^T & Q_1 + kpI \end{bmatrix} \begin{bmatrix} x_{n-1} \\ x_{n-1} \end{bmatrix} = \begin{bmatrix} \mu_{n-1}(Q_1)x_{n-1} + kpx_{n-1} - p\hat{B}x_{n-1} \\ \mu_{n-1}(Q_1)x_{n-1} + kpx_{n-1} - p\hat{B}^T x_{n-1} \end{bmatrix} \tag{8.14}$$

where $\hat{B}$ is a zero-one matrix satisfying $\hat{B} = \frac{B}{p}$. A circulant matrix is a matrix where each row is the same as the previous one, but the elements are shifted one position right and wrapped around at the end. Matrix B in (8.1) is an example of a circulant matrix. Circulant matrices are commutative [196]. If two matrices commute, the two matrices have the same set of eigenvectors [42]. When Q_1 and $\hat{B}$ are symmetric circulant matrices, Q_1 and $\hat{B}$ commute, i.e., $Q_1\hat{B} = \hat{B}Q_1$, and the eigenvectors of Q_1 and $\hat{B}$ are the same [42]. The eigenvector x_{n-1} of the Laplacian Q_1 is also an eigenvector of matrix $\hat{B}$ belonging to the eigenvalue λ, where $\lambda = \frac{x_{n-1}^T \hat{B} x_{n-1}}{x_{n-1}^T x_{n-1}} = 2x_{n-1}^T \hat{B} x_{n-1}$ because the normalization $x^T x = 2x_{n-1}^T x_{n-1} = 1$. Substituting $\hat{B}x_{n-1} = \lambda x_{n-1}$ in (8.14) yields

$$\begin{bmatrix} Q_1 + kpI & -p\hat{B} \\ -p\hat{B}^T & Q_1 + kpI \end{bmatrix} \begin{bmatrix} x_{n-1} \\ x_{n-1} \end{bmatrix} = \left(\mu_{n-1}(Q_1) + kp - \lambda p\right) \begin{bmatrix} x_{n-1} \\ x_{n-1} \end{bmatrix}$$

The vector $x = \left(x_{n-1}^T,\ x_{n-1}^T\right)^T$ is an eigenvector of Q belonging to eigenvalue $\mu = \mu_{n-1}(Q_1) + (k-\lambda)p$.

When the coupling weight p is small enough, the non-trivial eigenvalue $\mu^* = 2kp$ in (8.4) can be made the algebraic connectivity μ_{N-1} and eigenvalue $\mu_{n-1}(Q_1) + (k-\lambda)p$ can be made the third smallest eigenvalue μ_{N-2}. By increasing the coupling

weight p, a transition of the algebraic connectivity μ_{N-1} occurs, where $\mu^* = 2kp$ is no longer the second smallest one. The transition occurs at the point p^* such that $2kp^* = \mu_{n-1}(Q_1) + (k-\lambda)\,p^*$, from which

$$p^* = \frac{\mu_{n-1}}{k+\lambda}$$

where $\lambda = 2x_{n-1}^T \hat{B} x_{n-1}$.

Figure 8.6 shows the algebraic connectivity of the interdependent network consisting of two identical circulant graphs with a 2-to-2 ($k = 2$) interconnection. The size of each circulant graph is $n = 100$ with average degree $d_{av} = 6$. When the coupling strength $p \le p^*$, the algebraic connectivity μ_{N-1} is $4p$. When $p \ge p^*$, the algebraic connectivity in Figure 8.6 is analytically expressed as $\mu_{N-1} = \mu_{n-1}(Q_1) + (2-\lambda)p$. The transition occurs at the point $p^* = \frac{\mu_{n-1}}{2+\lambda}$, where $\lambda = 2x_{n-1}^T \hat{B} x_{n-1}$.

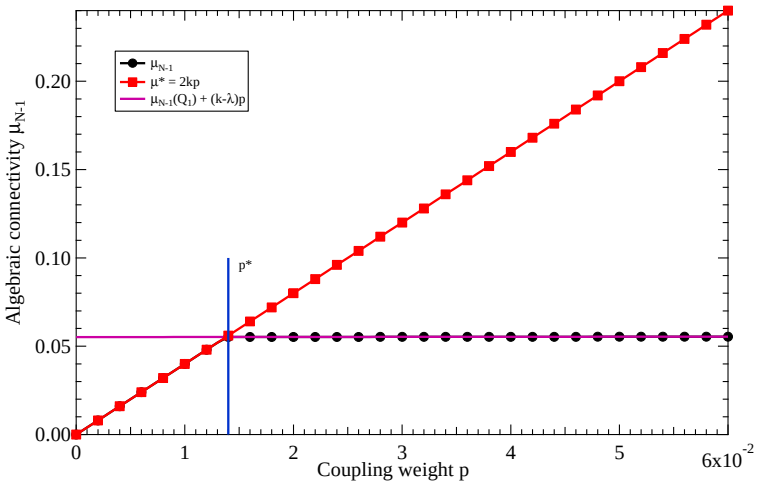


Figure 8.6: Algebraic connectivity in the interdependent network consisting of two identical circulant graphs with a 2-to-2 ($k = 2$) interconnection. The number of nodes for each circulant graph is $n = 100$ and the average degree is $d_{av} = 6$.

8.5.2. n-TO-n INTERCONNECTION

For an n-to-n interconnection pattern, the Laplacian matrix of the interdependent graph G reads

$$Q = \begin{bmatrix} Q_1 + pnI & -pJ_{n\times n} \\ -pJ_{n\times n} & Q_2 + pnI \end{bmatrix}$$

where $n \times n$ all-one matrix J represents that one node in graph G_1 connects to all nodes in graph G_2 and vice versa. Graph G is the join [193] of graphs G_1 and G_2 if the coupling weight $p = 1$.

Let x_1 be the eigenvector belonging to the eigenvalue $\mu_{n-1}(Q_1)$ of graph G_1 and x_2 be the eigenvector belonging to the eigenvalue $\mu_{n-1}(Q_2)$ of graph G_2. For vectors $x = \left(x_1^T,\ 0\right)^T$ and $x = \left(0,\ x_2^T\right)^T$, the eigenvalue equation for the Laplacian matrix Q of G can be written as

$$\begin{bmatrix} Q_1 + pnI & -pJ \\ -pJ & Q_2 + pnI \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = (\mu_{n-1}(Q_1) + np) \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} Q_1 + pnI & -pJ \\ -pJ & Q_2 + pnI \end{bmatrix} \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = (\mu_{n-1}(Q_2) + np) \begin{bmatrix} 0 \\ x_2 \end{bmatrix}$$

For an n-to-n ($k = n$) interconnection, the non-trivial eigenvalue $\mu^* = 2np$ can be made the algebraic connectivity $\mu_{N-1}(Q)$ of the Laplacian Q if the coupling weight p is small. The eigenvalue $\min\{\mu_{n-1}(Q_1),\ \mu_{n-1}(Q_2)\} + np$ can be made the third smallest eigenvalue $\mu_{N-2}(Q)$ with a small coupling weight p. The transition threshold p^* occurs when $\mu_{N-1}(Q) = \mu_{N-2}(Q)$ resulting in

$$p* = \min\left\{\frac{\mu_{n-1}(Q_1)}{n},\ \frac{\mu_{n-1}(Q_2)}{n}\right\} \tag{8.15}$$

Figure 8.7 shows the algebraic connectivity of the interdependent network consisting of two Erdős-Rényi graphs $G_p(n)$ with $n = 500$ nodes and average degree $d_{av} = 6$ and the interconnection pattern is n-to-n. Figure 8.7 demonstrates that when the coupling weight p is small, the algebraic connectivity is $\mu_{N-1} = 2np$. With the increase of p, the algebraic connectivity is described by $\mu_{N-1} = \min\{\mu_{n-1}(Q_1),\ \mu_{n-1}(Q_2)\} + np$. The transition occurs when $2np = \min\{\mu_{n-1}(Q_1),\ \mu_{n-1}(Q_2)\} + np$ and the threshold p^* obeys (8.15).

8.5.3. $(n-1)$-TO-$(n-1)$ INTERCONNECTION

When $B = p(J - I)$ and $G_2 = G_1$, the eigenvalue equation for the Laplacian matrix Q reads, with vector $x = \left(x_{n-1}^T,\ -x_{n-1}^T\right)^T$ where x_{n-1} is an eigenvector corresponding to the algebraic connectivity $\mu_{n-1}(Q_1)$ of graph G_1,

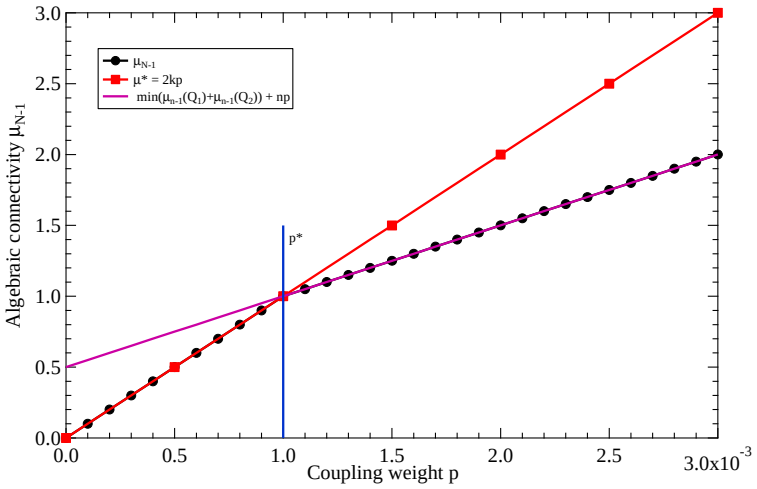


Figure 8.7: Algebraic connectivity μ_{N-1} in the interdependent network consisting of two Erdős-Rényi graphs with $n = 500$ nodes and average degree $d_{av} = 6$. The interconnection pattern is n-to-n.

$$\begin{bmatrix} Q_1 + p(n-1)I & -p(J-I) \\ -p(J-I) & Q_1 + p(n-1)I \end{bmatrix} \begin{bmatrix} x_{n-1} \\ -x_{n-1} \end{bmatrix} = \left(\mu_{n-1}(Q_1) + (n-2)\,p\right) \begin{bmatrix} x_{n-1} \\ -x_{n-1} \end{bmatrix} \tag{8.16}$$

The non-trivial eigenvalue follows $\mu^* = 2(n-1)p$ for an $(n-1)$-to-$(n-1)$ interconnection. When p is small, the eigenvalue $2(n-1)p$ can be made the algebraic connectivity μ_{N-1} and the eigenvalue $\mu_{n-1}(Q_1) + (n-2)\,p$ can be made the third smallest eigenvalue μ_{N-2}. At the transition threshold p^*, we have that $\mu_{N-1} = \mu_{N-2}$ from which the threshold p^* follows

$$p* = \frac{\mu_{n-1}(Q_1)}{n}$$

8.5.4. A GRAPH COUPLED WITH ITS COMPLEMENTARY GRAPH

The complementary graph G_1^c of a graph G_1 has the same set of nodes as G_1 and two nodes are connected in G_1^c if they are not connected in G_1 and vice versa [42]. The adjacency matrix of the complementary graph G_1^c is $A_1^c = J - I - A_1$. The Laplacian of the complement graph G_1^c follows $nI - J - Q_1$.

For an interdependent graph G consisting of a graph G_1 and its complementary graph G_1^c with an n-to-n interconnection pattern, the Laplacian matrix Q of the

interdependent graph G reads

$$Q = \begin{bmatrix} Q_1 + npI & -pJ \\ -pJ & nI - J - Q_1 + npI \end{bmatrix}$$

Let x_{n-1} be the eigenvector belonging to the eigenvalue μ_{n-1} of the graph G_1 and x_1 be the eigenvector belonging to the eigenvalue μ_1. For vectors $x = \left(x_{n-1}^T,\ 0\right)^T$ and $x = \left(0,\ x_1^T\right)^T$, the eigenvalue equation for the Laplacian matrix Q of G can be written as

$$\begin{bmatrix} Q_1 + npI & -pJ \\ -pJ & nI - J - Q_1 + npI \end{bmatrix} \begin{bmatrix} x_{n-1} \\ 0 \end{bmatrix} = (\mu_{n-1}(Q_1) + np) \begin{bmatrix} x_{n-1} \\ 0 \end{bmatrix} \tag{8.17}$$

$$\begin{bmatrix} Q_1 + npI & -pJ \\ -pJ & nI - J - Q_1 + npI \end{bmatrix} \begin{bmatrix} 0 \\ x_1 \end{bmatrix} = (n + np - \mu_1(Q_1)) \begin{bmatrix} 0 \\ x_1 \end{bmatrix} \tag{8.18}$$

When the coupling weight p is small, eigenvalue $\mu^* = 2np$ can be made the algebraic connectivity $\mu_{N-1}(Q) = \mu^* = 2np$ and eigenvalue $\min\left(\mu_{n-1}(Q_1) + np,\ n + np - \mu_1(Q_1)\right)$ can be made the third smallest eigenvalue $\mu_{N-2}(Q)$. From $\mu_{N-1}(Q) = \mu_{N-2}(Q)$ at the transition point p^*, we arrive at

$$p* = \min\left(\frac{\mu_{n-1}(Q_1)}{n},\ 1 - \frac{\mu_1(Q_1)}{n}\right)$$

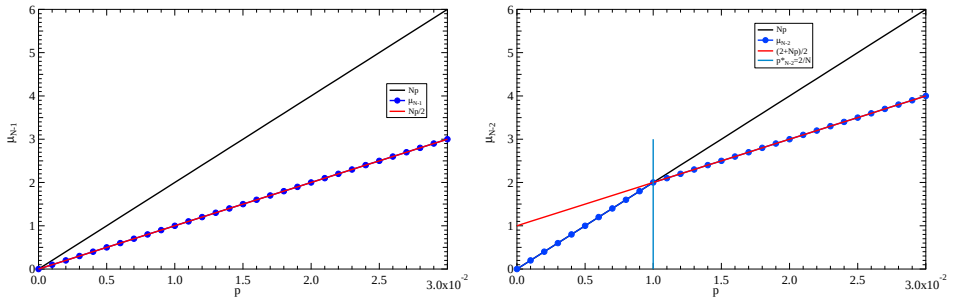
8.5.5. AN EXAMPLE OF THE NON-EXISTENCE OF THE STRUCTURAL TRANSITION

In this subsection, we consider an interdependent network consisting of a star graph G_1 and its complementary graph G_1^c while the interconnection pattern is n-to-n. For a star graph with size n, the eigenvalues of the Laplacian [42] are 0, 1 with multiplicity $n-2$ and n. Substituting $\mu_{n-1}(Q_1) = 1$ and $\mu_1(Q_1) = n$ into eigenvalue equations (8.17) and (8.18) yields two eigenvalues np and $np+1$.

When the coupling weight $p > 0$, the non-trivial eigenvalue $\mu^* = 2np$ cannot be the second smallest eigenvalue of the Laplacian Q because it is always larger than the eigenvalue np. Hence, the transition between μ^* and the algebraic connectivity $\mu_{N-1}(Q)$ will never occur as shown in Figure 8.8(a). Instead, when p is small, the non-trivial eigenvalue $\mu^* = 2np$ can be made the third smallest eigenvalue $\mu_{N-2}(Q)$. By increasing the coupling weight p, the eigenvalue $\mu^* = 2np$ may no longer be the

third smallest eigenvalue of the Laplacian Q. There exists a threshold denoted as p^*_{N-2} such that $\mu^* = 2np$ exceeds $\mu_{N-2}(Q)$ when $p > p^*_{N-2}$.

When $p \leq p^*_{N-2}$ then the third smallest eigenvalue follows $\mu_{N-2}(Q) = 2np$. Above the transition point p^*_{N-2}, the non-trivial eigenvalue $\mu^* = 2np$ exceeds eigenvalue $1 + np$. The transition occurs when $2np* = 1 + np*$ resulting in $p^*_{N-2} = \frac{1}{n}$. Figure 8.8(b) shows that the transition occurs at the point $p^*_{N-2} = \frac{1}{n}$.



(a) Algebraic connectivity μ_{N-1} (b) Third smallest eigenvalue μ_{N-2}

Figure 8.8: An interdependent network consisting of a star graph and its complement with n-to-n interdependency. The size n of the star graph is 100.

In the above example, the complementary graph G_1^c of a star is a disconnected graph. The hub node in the star G_1 is an isolated node in graph G_1^c. The coupling is stronger between graph G_1 and the connected component in graph G_1^c than that between graph G_1 and the isolated node in G_1^c. The isolated node first decouples from the interdependent network G before the connected component in G_1^c decouples from interdependent graph G. As a result, the structural transition in p occurs at the third smallest eigenvalue rather than at the second smallest eigenvalue. The above example also agree with the upper bound in (8.9) that the threshold $p^* = 0$ when $\mu_{n-1}(Q_1) = 0$ or $\mu_{n-1}(Q_2) = 0$. There is no transition between non-trivial eigenvalue $\mu^* = 2kp$ and the algebraic connectivity μ_{N-1}, if one of the coupled graphs is disconnected.

8.6. CONCLUSION

We generalize a one-to-one interconnection to a general k-to-k interconnection for interdependent networks. The interconnection matrix B representing the

k-to-k interconnection obeys regularity (constant row and column sum) and a non-trivial eigenvalue of such interdependent networks can be deduced [178]. For $B = pI$ (one-to-one interconnection), it has been shown [182, 184] that there exists a structural transition p^*: when $p < p^*$, the network acts as separated graphs G_1 and G_2; when $p > p^*$, the network acts as a whole.

For a general k-to-k interconnection ($B \neq pI$ unless $k = 1$), we analyse the properties of the transition threshold p^*. For connected graphs G_1 and G_2, we show that the transition threshold p^* is upper bounded by the minimum algebraic connectivity of graphs G_1 and G_2 divided by k for a k-to-k interconnection. If graph G_1 is the join of subgraphs each with k nodes and the matrix B is singular with the kernel vector to be the eigenvector of the algebraic connectivity of graph G_1 (as shown in (8.10) and (8.12)), then the upper bound for the transition threshold p^* is attained. The upper bounds and the exact value of the transition threshold p^* can be applied for the identification of the interaction and multi-layer coupling pattern of neural networks given that a healthy human brain operates around the transition point [187].

In addition, the physical meaning of the threshold p^* is that below the transition threshold p^*, the minimum cut of the network includes all the interconnection links, whereas above the transition threshold p^*, the minimum cut contains both the interconnection links between graphs G_1 and G_2 and the links within G_1 and G_2. For special topologies (as specified in Section 8.5), the threshold p^* can be determined exactly. If one of the graphs G_1 or G_2 is disconnected, then the structural threshold p^* for the algebraic connectivity does not exist.

9

CONCLUSION

The thesis investigates the robustness of complex networks including theoretical approaches and application of theories to real-world networks. The main contribution of the thesis is a better understanding on how topologies and properties of networks influence the structural and dynamical robustness of networks. The take-away message of this thesis is that the analysis of the robustness of networks is a multi-objective problem. Failure scenarios, quantification or characterization of the robustness anticipating the specific functionality of a specific network, topological and dynamical properties networks and the complex nature of real-world networks all play a role in designing a robust network or enhancing an existing network.

9.1. MAIN CONTRIBUTIONS

Chapter 2 contributes to the theoretical and experimental findings that are applicable in real-world scenarios such as single-line instalments in infrastructural networks or single-line protection against cyber-physical attacks. The upper and lower bounds introduced in chapter 2 can be used to support policy and decision makers to choose a line to install or protect given certain operational costs. Moreover, when computational cost for finding optimal links to add or remove is pro-

hibitive, the topological and spectral strategies can still indicate links resulting in a high robustness level. If the optimal added or removed links for the algebraic connectivity are known, then the respective links for the effective graph resistance are different but in close proximity.

Chapter 3 employs the effective graph resistance as a robustness metric for network expansions to improve the grid robustness against cascading failures. The effective graph resistance takes the multiple paths and their ability to accommodate power flows into account to quantify the robustness of power grids. The experimental verification on IEEE power systems demonstrates the effectiveness of the effective graph resistance to identify single links that improve the grid robustness against cascading failures. Additionally, when computational cost for finding optimal links is prohibitive, strategies that optimize the effective graph resistance can still identify an added link resulting in a high level of robustness. The occurrence of Braess's paradox in power grids suggests that the robustness can be occasionally decreased by placing additional links. In particular, a badly designed power grid may cause enormous costs for new lines that actually reduce the grid robustness.

Chapter 4 investigates the robustness of metro networks by analysing ten theoretical robustness metrics and three numerical metrics. For the latter, we investigate two critical thresholds f, when 90% of the network is still remaining, $f_{90\%}$ (both under random failure and targeted attack), and when the complete network is disintegrated, f_c (under targeted attack). We find that the ten theoretical robustness metrics capture two distinct aspects of the robustness of metro networks. A first aspect deals with the number of alternative paths, suggesting that more alternative paths are more desirable. In contrast, the second aspect deals with "resistance", suggesting that longer lines with no shorter alternative paths perform poorly. As metro networks are expanded, effort should be put into creating transfer stations, both in city centres and peripheral areas to ensure that not only many alternative paths are created to reach a destination, but also that the average number of stations between two transfers is kept to a minimum. To fully capture these two aspects and assess the robustness of metro networks, we plot the ten theoretical measures (standardized) on radar plots. This method offers both an equal representation of the variables at play as well as an aesthetically-pleasing visual aid to help planners in their task to design robust metro networks.

Chapter 5 investigates fundamental properties including the degree distribution and the assortativity of line graphs, which transform links from a graph to nodes in its line graph. The degree distribution of the line graph of an Erdős-Rényi random graph follows the same pattern of the degree distribution as the original graph. We derive a formula for the assortativity of a line graph which indicates that the assortativity of a line graph is not linearly related to the assortativity of its original graph. Moreover, the assortativity is positive for the line graphs of Erdős-Rényi graphs, Barabási-Albert graphs and most real-world networks. In contrast, certain types of trees, path and path-like graphs, have negative assortativity in their line graphs. Furthermore, non-trees consisting of cycles and paths can also have negative assortativity in their line graphs.

Chapter 6 studies the eigenvector matrix Z of the Laplacian matrix Q for a graph G with the aim to understand how properties of matrix Z contain information about the structure of graph G. We find that the sum s_Z of all the elements in Z increases with the size of the graph as $O(\sqrt{N})$. The higher the average degree in a graph, the lower the number of zeros in the eigenvector matrix. Moreover, the distribution of the maximum element in the eigenvector matrix is the same as the distribution of the minimum element. The row sum of the eigenvector matrix Z of the Laplacian Q, coined the dual fundamental weight φ, in Erdős-Rényi random graphs follows closely the product of a Gaussian and a super-Gaussian distribution.

Chapter 7 investigates two interconnection topologies for interdependent networks that incorporate the locations of nodes. The two topologies generalize the one-to-one interconnection to an arbitrary number of interconnections depending on the locations of nodes. We analyse the properties of the two topologies and the impact of the two interconnection topologies on robustness of interdependent networks against cascading failures. Specifically, the derivation of the number of links in the two topologies enables the comparison of robustness performance between the two topologies. We fined the random geometric graph provides the flexibility for network providers to determine the link density of interconnections in order to achieve the desired robustness level. The relative neighbourhood graph, often used in wireless networks [177] to provide optimal coverage with least energy consumption, as an interconnection topology is less robust compared to the random geometric graph.

In addition, we propose the derivative of the largest mutually connected component as a new robust metric which addresses the impact of a small fraction of failed nodes. To avoid the collapse of the whole network, the proposed robustness metric quantifies the damage of networks triggered by a small fraction of failures, significantly smaller than the fraction at the critical threshold, that corresponds to the collapse of the whole network.

Chapter 8 investigates the structural transition threshold for the interdependent network consisting of two graphs G_1 and G_2 with a regular interconnection pattern. The transition threshold p^* is upper bounded by the minimum algebraic connectivity of graphs G_1 and G_2 divided by k for a k-to-k interconnection. The upper bound for the transition threshold p^* is attained if graph G_1 is the join of subgraphs each with k nodes and the matrix B is singular with the kernel vector to be the eigenvector of the algebraic connectivity of graph G_1. The physical meaning of the threshold p^* is that below the transition threshold p^*, the minimum cut of the network includes all the interconnection links, whereas above the transition threshold p^*, the minimum cut contains both the interconnection links between graphs G_1 and G_2 and the links within G_1 and G_2. For special topologies, the threshold p^* can be determined exactly. If one of the graphs G_1 or G_2 is disconnected, then the structural threshold p^* for the algebraic connectivity does not exist.

9.2. DIRECTIONS FOR FUTURE WORK

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

1. The influence of the topology of a network on the dynamic process in that network is intensively explored in complex networks. However, the inverse, how the dynamic process influences the topology of a growing network, is rarely studied. It would be interesting to investigate the influence of dynamic processes on the topology of a network. A real-world example is that the traffic flow in a transportation network impacts the extension of that transportation network.

2. In chapter 2, we explore adding single links into an existing network or removing single links from that network. The goal is to determine the link whose ad-

dition maximally increases the robust and the link to protect under a limited budget. A generalized question, which remains open, is that the determination of multiple links whose addition maximally increases the robustness of a network.

3. Chapter 7 investigates the robustness of interdependent networks under random failures. However, targeted attacks, as happened in most real-world networks, might severely destroy the network. Which node or which set of nodes will dramatically destroy the network upon removal? Safely protecting such nodes results in a high level of the robustness of interdependent networks.

4. Power networks are subject to failures of transmission lines. The line graph, studied in chapter 5, transforms links in the original graph into nodes. The study of failures of nodes in line graphs enables a better understanding on the impact of link failures on power grids. The relation between the impact of failures of links in the original graph and the impact of failures of the corresponding nodes in the line graph, remains open.

5. Spectral metrics, such as spectral radius, algebraic connectivity, play a key role in characterizing network robustness. However, the eigenvectors of graph matrices are rarely explored. The results in chapter 6 are only a tip of the iceberg. Various questions remain open. For example, how the dual fundamental weight (row sum of the eigenvector matrix of the Laplacian) relates to the properties, i.e., importance of a node, of a graph? Does the dual fundamental weight provide a better graph partition than the Fiedler vector? Is there a correlation between the dual fundamental weight and the degree vector of a graph?

A

BOUNDS FOR THE ALGEBRAIC CONNECTIVITY

In this appendix, we derive upper and lower bounds for the algebraic connectivity in terms of the effective graph resistance.

The analogy of inequality (2.8) is:

$$\frac{N-2}{\sum_{j=1}^{N-2}\frac{1}{\mu_j}} \leq \frac{1}{N-2}\sum_{j=1}^{N-2}\mu_j$$

Introducing the definition $S = \sum_{j=1}^{N-2}\frac{1}{\mu_j}$, with the sum of all the eigenvalues [42] satisfying $\sum_{j=1}^{N-1}\mu_j = 2L$, it follows that

$$\frac{N-2}{S} \leq \frac{2L-\mu_{N-1}}{N-2}$$

With the definitions $S = \frac{R_G}{N} - \frac{1}{\mu_{N-1}}$, $\alpha_G = \mu_{N-1}$ and by assuming a connected graph ($\mu_{N-1} > 0$), it holds, for $N > 2$

$$\alpha_G \leq 2L - \frac{(N-2)^2}{\frac{R_G}{N} - \frac{1}{\alpha_G}}$$

which is transformed into a quadratic inequality of α_G:

$$\frac{R_G}{N}\alpha_G^2 + ((N-2)^2 - 1 - 2L\frac{R_G}{N})\alpha_G + 2L \leq 0 \tag{A.1}$$

In a factored form and by denoting $\frac{2LR_G}{N} = \widetilde{R_G}$, the quadratic inequality (A.1) is expressed as follows:

$$0 \geq \left(\alpha_G - \frac{\widetilde{R_G} - (N-1)(N-3) - \xi}{\widetilde{R_G}/L}\right)\left(\alpha_G - \frac{\widetilde{R_G} - (N-1)(N-3) + \xi}{\widetilde{R_G}/L}\right) \tag{A.2}$$

where $\xi = \sqrt{[\widetilde{R_G} - (N-3)^2][\widetilde{R_G} - (N-1)^2]}$ is the squareroot of the discriminant. The lower [42] bound $R_G \geq \frac{(N-1)^2}{E[D]}$, rephrased as $\widetilde{R_G} \geq (N-1)^2$, shows that $\widetilde{R_G} - (N-3)^2 > 0$ and $\widetilde{R_G} - (N-1)(N-3) > 0$, hence, ξ is real. Therefore, the quadratic equation in (A.1) has the following two real roots:

$$\begin{aligned} x_1 &= \frac{\widetilde{R_G} - (N-1)(N-3) - \xi}{\widetilde{R_G}/L} \\ x_2 &= \frac{\widetilde{R_G} - (N-1)(N-3) + \xi}{\widetilde{R_G}/L} \end{aligned}$$

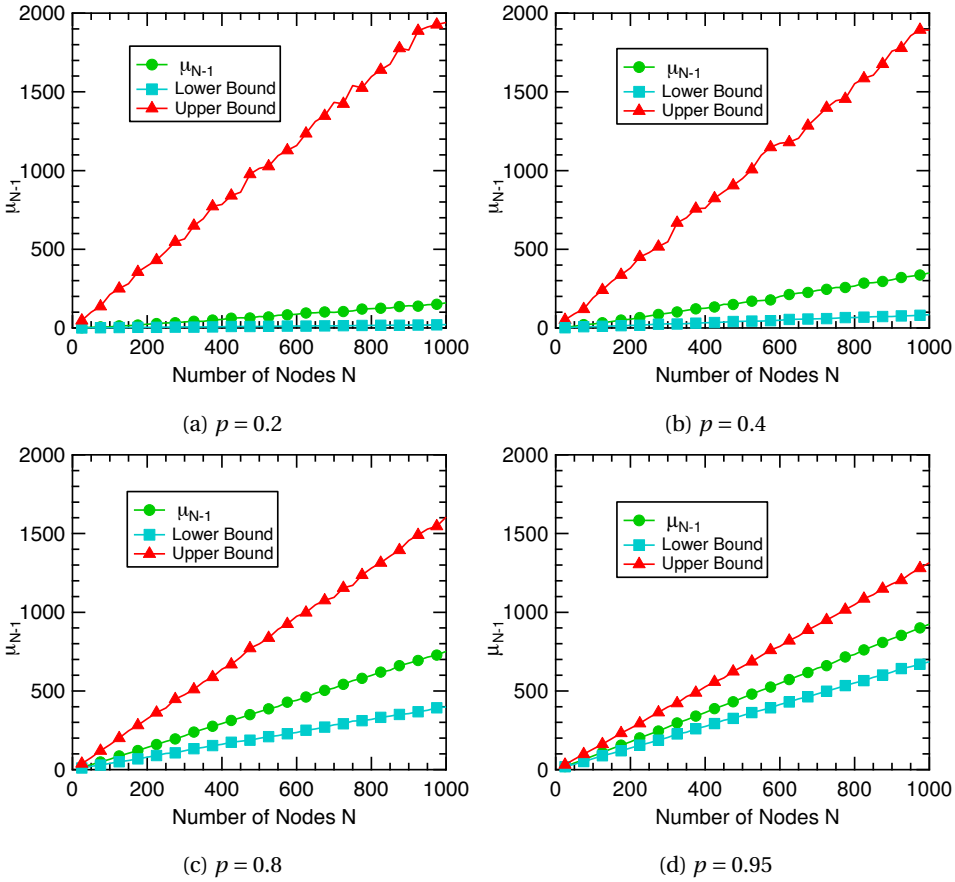
Vieta's formula indicates that the product of roots equals $x_1 x_2 = \frac{2L}{\frac{R_G}{N}} > 0$ that results in both x_1 and x_2 being either positive or negative. Since $x_2 > 0$, the root x_1 is also positive. In summary, we deduce a new lower bound:

$$\alpha_G \geq L\left(1 - \frac{(N-1)(N-3)}{\widetilde{R_G}} - \sqrt{[1 - \frac{(N-3)^2}{\widetilde{R_G}}][1 - \frac{(N-1)^2}{\widetilde{R_G}}]}\right)$$

and an upper bound for the algebraic connectivity:

$$\alpha_G \leq L\left(1 - \frac{(N-1)(N-3)}{\widetilde{R_G}} + \sqrt{[1 - \frac{(N-3)^2}{\widetilde{R_G}}][1 - \frac{(N-1)^2}{\widetilde{R_G}}]}\right)$$

Figure A.1 illustrates the lower and upper bounds of the algebraic connectivity α_G for Erdős-Rényi graphs with different link density p. As link density increases, the upper and lower bounds come closer. The bounds converge to the algebraic connectivity resulting in an equality for (15).



(a) $p = 0.2$

(b) $p = 0.4$

(c) $p = 0.8$

(d) $p = 0.95$

Figure A.1: Upper and lower bounds of the algebraic connectivity α_G.

B

PROOFS FOR LINE GRAPHS

This appendix presents three proofs for the formulas in chapter 5.

B.1. PROOF OF EQUATION (5.6)

The degree distribution in scale free graphs G is

$$\Pr[D = k] = \frac{k^{-\gamma}}{c_1}, \quad k = s, \cdots, K \tag{B.1}$$

where $c_1 = \sum_{k=s}^{K} k^{-\gamma}$ is the normalization constant and s is the minimum degree and K is the maximum degree in G. Assuming the node degrees in the scale free graph are independent, the generating function for the line graph of scale free graphs can be written as equation (5.5). Substituting the derivative of the generating function $\varphi_D^{'}(z) = \frac{1}{E[D]} \sum_{k=0}^{N-1} k z^{k-1} \Pr[D = k]$ and the average degree $E[D] = \sum_{k=0}^{N-1} k \Pr[D = k] = \frac{c_2}{c_1}$, where $c_2 = \sum_{k=s}^{K} k^{1-\gamma}$, into equation (5.5) yields

$$\varphi_{D_l}(z) = \left(\frac{c_1}{c_2}\right)^2 \left(\varphi_D^{'}(z)\right)^2 \tag{B.2}$$

and the Taylor coefficients obey

$$\Pr[D_l = k] = \left(\frac{c_1}{c_2}\right)^2 \frac{1}{k!} \frac{d^k \left(\varphi_D^{'}(z)\right)^2}{dz^k}\bigg|_{z=0}$$

B

Using the Leibniz's rule $(fg)^{(k)} = \sum_{m=0}^{k} \binom{k}{m} f^{(m)} g^{(k-m)}$, where $f = g = \varphi_D^{'}(z)$, yields

$$\Pr[D_l = k] = \left(\frac{c_1}{c_2}\right)^2 \frac{1}{k!} \sum_{m=0}^{k} \binom{k}{m} \frac{d^{m+1}(\varphi_D(z))}{dz^{m+1}} \frac{d^{k-m+1}(\varphi_D(z))}{dz^{k-m+1}}\bigg|_{z=0}$$

Substituting $k!\Pr[D = k] = \frac{d^k(\varphi_D(z))}{dz^k}\Big|_{z=0}$, we arrive at

$$\Pr[D_l = k] = \frac{1}{k!}\left(\frac{c_1}{c_2}\right)^2 \sum_{m=0}^{k} \frac{k!}{m!(k-m)!}(m+1)!\Pr[D = m+1](k-m+1)!\Pr[D = k-m+1]$$

Applying the power law degree distribution in equation (B.1), we have

$$\Pr[D_l = k] = \frac{1}{c_2^2} \sum_{m=1}^{k+1} \Big(m(k+2-m)\Big)^{1-\gamma} \tag{B.3}$$

For $\gamma = 3$, we transform equation (B.3) in the following form:

$$c_2^2 \Pr[D_l = k] = \frac{1}{(k+2)^3} \sum_{i=1}^{k+1} \frac{1}{(\frac{i}{k+2})^2(1-\frac{i}{k+2})^2} \frac{1}{k+2} \tag{B.4}$$

We use the following expression between a sum in the limit to infinity and a definite integral [197]

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\Delta x$$

We set $\Delta x = \frac{1}{k+2}$, $x_i = i\Delta x = \frac{i}{k+2}$, $f(x) = \frac{1}{x^2(1-x)^2}$ and (B.4) boils down to

$$c_2^2 \Pr[D_l = k] = \frac{1}{(k+2)^3} \sum_{i=1}^{k+1} f(x_i)\Delta x \tag{B.5}$$

We consider the case of limit to infinity for k ($k \to \infty$) or k very large and evaluate the sum $\sum_{i=1}^{k+1} f(x_i)\Delta x$, which can be transformed into

$$\sum_{i=1}^{k+1} f(x_i)\Delta x \approx \int_{\frac{1}{k+2}}^{\frac{k+1}{k+2}} f(x)dx \tag{B.6}$$

Now,

$$\int_{\frac{1}{k+2}}^{\frac{k+1}{k+2}} f(x)dx = \int_{\frac{1}{k+2}}^{\frac{k+1}{k+2}} \frac{1}{x^2(1-x)^2}dx$$
$$= \int_{\frac{1}{k+2}}^{\frac{k+1}{k+2}} \left(\frac{2}{x} + \frac{2}{1-x} + \frac{1}{x^2} + \frac{1}{(1-x)^2}\right)dx$$
$$= 2\left(2\ln(k+1) + \frac{k(k+2)}{k+1}\right) \tag{B.7}$$

Using (B.7) and (B.6) into (B.5), leads to

$$c_2^2 \Pr[D_l = k] \approx \frac{2}{(k+2)^2}\left(\frac{2\ln(k+1)}{k+2} + \frac{k}{k+1}\right)$$

Since $\lim_{k\to\infty} \frac{\ln(k+1)}{k+2} = 0$ and $\lim_{k\to\infty} \frac{k}{k+1} = 1$, we arrive at

$$\Pr[D_l = k] \approx \frac{1}{c_2^2}(k+2)^{-2} \tag{B.8}$$

B.2. PROOF FOR THEOREM 6

Proof. A link l with end nodes l^+ and l^- in the line graph $l(G)$ corresponds to a connected triplet in G. Without loss of generality, we assume that nodes l^+ and l^- in the line graph correspond to links $l^+ = i \sim c$ and $l^- = j \sim c$, where links $i \sim c$ and $j \sim c$ share a common node c, in the original graph as shown in Figure B.1. The degree in line graph is $d_{l^+} = d_i + d_c - 2$ and $d_{l^-} = d_j + d_c - 2$. Since subtracting

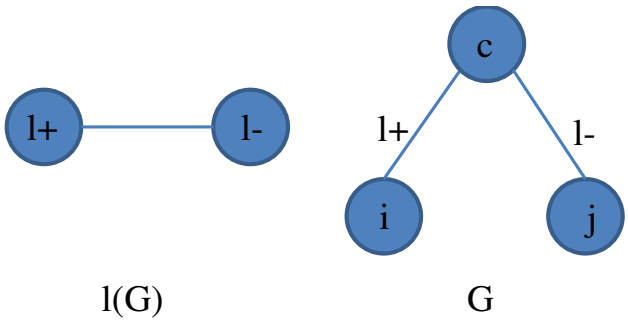


Figure B.1: Link transformation.

2 everywhere does not change the linear correlation coefficient, we proceed with $d_{l^+} = d_i + d_c$ and $d_{l^-} = d_j + d_c$. First, we compute the joint expectation

$$
\begin{aligned}
E[D_{l^+} D_{l^-}] &= \frac{\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{c=1}^{N} (d_i + d_c)(d_j + d_c) a_{ic} a_{jc}}{2L_{l(G)}} \\
&= \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{c=1}^{N} (d_i + d_c)(d_j + d_c) a_{ic} a_{jc} - \sum_{i=1}^{N} \sum_{c=1}^{N} (d_i + d_c)^2 a_{ic}^2}{2L_{l(G)}} \\
&= \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{c=1}^{N} d_i a_{ic} a_{jc} d_j + 2 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{c=1}^{N} d_i a_{ic} a_{jc} d_c + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{c=1}^{N} d_c^2 a_{ic} a_{jc}}{2L_{l(G)}} \\
&\quad - \frac{2 \sum_{i=1}^{N} \sum_{c=1}^{N} d_i^2 a_{ic}^2 + 2 \sum_{i=1}^{N} \sum_{c=1}^{N} d_i a_{ic}^2 d_c}{2L_{l(G)}}
\end{aligned}
$$

With $\sum_{j=1}^{N} a_{jc} = d_c$, we arrive at

$$
E[D_{l^+} D_{l^-}] = \frac{d^T A^2 d + 2 d^T A \Delta d + \sum_{i=c}^{N} d_c^4 - 2 \sum_{i=1}^{N} d_i^3 - 2 d^T A d}{2L_{l(G)}} \tag{B.9}
$$

The average degree $E[D_{l^+}] = E[D_i + D_c]$ is the average degree of two connected nodes i and c from a triplet (see Figure B.1) in the original graph. Thus,

$$
\begin{aligned}
E[D_{l^+}] &= \frac{\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{c=1}^{N} (d_i + d_c) a_{ic} a_{jc}}{2L_{l(G)}} \\
&= \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{c=1}^{N} d_i a_{ic} a_{jc} + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{c=1}^{N} d_c a_{ic} a_{jc} - \sum_{i=1}^{N} \sum_{c=1}^{N} d_i a_{ic}^2 - \sum_{i=1}^{N} \sum_{c=1}^{N} d_c a_{ic}^2}{2L_{l(G)}}
\end{aligned}
$$

from which

$$
E[D_{l^+}] = \frac{d^T A d + \sum_{c=1}^{N} d_c^3 - 2 d^T d}{2L_{l(G)}} \tag{B.10}
$$

The variance $\sigma^2_{D_{l^+}} = Var[D_{l^+}] = E[D^2_{l^+}] - (E[D_{l^+}])^2$ and

$$\begin{aligned} E[D^2_{l^+}] &= \frac{\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{c=1}^{N}(d_i+d_c)^2 a_{ic}a_{jc}}{2L_{l(G)}} \\ &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{c=1}^{N} d_i^2 a_{ic}a_{jc} + 2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{c=1}^{N} d_i a_{ic}a_{jc}d_c + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{c=1}^{N} d_c^2 a_{ic}a_{jc}}{2L_{l(G)}} \\ &\quad - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{N} d_i^2 a_{ic}^2 + 2\sum_{i=1}^{N}\sum_{c=1}^{N} d_i a_{ic}^2 d_c}{2L_{l(G)}} \end{aligned}$$

which we rewrite as

$$E[D^2_{l^+}] = \frac{3d^T A\Delta d + \sum_{c=1}^{N} d_c^4 - 2\sum_{i=1}^{N} d_i^3 - 2d^T Ad}{2L_{l(G)}} \tag{B.11}$$

The number of links $L_{l(G)}$ in a line graph is [42]

$$L_{l(G)} = \frac{1}{2}d^T d - L = \frac{1}{2}(N_2 - N_1) \tag{B.12}$$

After substituting equations (B.9-B.12) into (5.7), we establish the Theorem. □

B.3. PROOF FOR COROLLARY 1

Proof. Using the variance $\sigma^2_{D_{l^+}} = \mathrm{Var}[D_{l^+}] = E[D^2_{l^+}] - (E[D_{l^+}])^2$, we rewrite the definition of assortativity (5.7) as

$$\rho_{D_{l(G)}} = 1 + \frac{E[D_{l^+}D_{l^-}] - E[D^2_{l^+}]}{\sigma^2_{D_{l^+}}} \tag{B.13}$$

According to equations (B.9) and (B.11), we have that

$$\begin{aligned} E[D_{l^+}D_{l^-}] - E[D^2_{l^+}] &= \frac{d^T A^2 d - d^T A\Delta d}{2L_{l(G)}} \\ &= \frac{N_4 - d^T A\Delta d}{2L_{l(G)}} \end{aligned} \tag{B.14}$$

The variance $\mathrm{Var}[D_{l^+}]$ of the end node of an arbitrarily chosen link can be written in terms of the variance $\mathrm{Var}[D_{l(G)}]$ of an arbitrarily chosen node [198]

$$\sigma^2_{D_{l^+}} = \frac{\mu u_3 - (\mathrm{Var}[D_{l(G)}])^2 + \mu^2 \mathrm{Var}[D_{l(G)}]}{\mu^2} \tag{B.15}$$

where $\mu = E[D_{l(G)}]$ and $u_3 = E[(D_{l(G)} - E[D_{l(G)}])^3]$. The variance $\mathrm{Var}[D_{l(G)}]$ of an arbitrarily chosen node can be written in terms of the assortativity [42]

$$\mathrm{Var}[D_{l(G)}] = 2(1+\rho_D)\left(\frac{1}{N_1}\sum_{i=1}^{N} d_i^3 - \left(\frac{N_2}{N_1}\right)^2\right) \tag{B.16}$$

Substituting (B.14-B.16) into (B.13), we prove the Corollary 1. □

C

PROPERTY OF THE EIGENVECTOR MATRIX OF THE LAPLACIAN FOR A DISCONNECTED GRAPH

In this appendix, we deduce the sum s_Z of elements in the eigenvector matrix Z of the Laplacian Q for a disconnected graph. We write the $N \times N$ symmetric matrix A as a block matrix

$$A = \begin{bmatrix} A_1 & B \\ B^T & A_2 \end{bmatrix}$$

where A_1 is an $(N-m) \times (N-m)$ symmetric matrix and A_2 is a $m \times m$ symmetric matrix with[1] $0 \leq m < \frac{N}{2}$. For example, for a graph G, A_1 and A_2 are the adjacency matrices of two subgraphs G_1 and G_2 of G, B represents the interconnection matrix of the links between G_1 and G_2. The eigenvalue equation $Ax = \lambda(A)\, x$ is written as the linear block set, with the eigenvector $x^T = \begin{bmatrix} v_{(N-m)\times 1} & w_{m\times 1} \end{bmatrix}^T$,

$$\begin{cases} A_1 v + B w = \lambda(A)\, v \\ B^T v + A_2 w = \lambda(A)\, w \end{cases}$$

[1] If $m \geq \frac{N}{2}$, we can interchange subgraph G_1 and G_2 so that $m < \frac{N}{2}$.

where we choose the normalization $x^T x = 1$, equivalent to $v^T v + w^T w = 1$. If the coupling matrix $B = 0$, then the set simplifies to

$$\begin{cases} A_1 v = \lambda(A)\, v \\ A_2 w = \lambda(A)\, w \end{cases}$$

which illustrates that v and w are eigenvectors (satisfying $v^T v + w^T w = 1$) belonging to the eigenvalue $\lambda(A)$, which is also an eigenvalue of at least one matrix, A_1 or A_2, because an eigenvector x is different from the zero vector, so that not both v and w can be the zero vector.

In the case of the Laplacian Q of G, where u is an eigenvector of Q_1, Q_2 and Q belonging to eigenvalue $\mu = 0$, then it holds that

$$\begin{cases} Q_1 v = 0 \\ Q_2 w = 0 \end{cases}$$

where $v = \alpha u$ and $w = \beta u$ with $1 = \alpha^2 (N - m) + \beta^2 m$. The latter is the equation of an ellipse with the two main axes $\frac{1}{\sqrt{N-m}}$ and $\frac{1}{\sqrt{m}}$,

$$\frac{\alpha^2}{\left(\frac{1}{\sqrt{N-m}}\right)^2} + \frac{\beta^2}{\left(\frac{1}{\sqrt{m}}\right)^2} = 1 \tag{C.1}$$

and any set (α, β) with both $\alpha \neq 0$ and $\beta \neq 0$ on the ellipse is a solution. Hence[2], for $m > 0$, there exists infinitely many normalizations of the eigenvector of Q belonging to the eigenvalue $\mu_N = 0$. When $m \to 0$ (and hence $\beta = 0$), the ellipse degenerates into the points $\alpha = \pm\frac{1}{\sqrt{N}}$. Moreover, we can construct two orthogonal eigenvectors (since the multiplicity of $\mu = 0$ is two). Let $x_1^T = \begin{bmatrix} \alpha u & \beta u \end{bmatrix}^T$ and $x_2^T = \begin{bmatrix} \gamma u & \delta u \end{bmatrix}^T$, where (γ, δ) is also a point on the above ellipse. Orthogonality requires that

$$0 = x_1^T x_2 = \begin{bmatrix} \alpha u & \beta u \end{bmatrix}^T \begin{bmatrix} \gamma u \\ \delta u \end{bmatrix} = \alpha\gamma (N - m) + \beta\delta m$$

leading to

$$\gamma = -\frac{\beta m}{\alpha (N - m)}\delta$$

[2]When there are c disconnected subgraphs in G, the normalization procedure results in c-dimensional ellipsoid leading to $c - 1$ degrees of freedom to normalize the c eigenvectors belonging to eigenvalue $\mu_N = 0$ of Q.

but also $1=\gamma^2(N-m)+\delta^2 m$. Combined yields $\delta=\pm\frac{1}{\sqrt{\left(\frac{\beta m}{\alpha\sqrt{N-m}}\right)^2+m}}$ and, after using $1=\alpha^2(N-m)+\beta^2 m$, we find

$$\delta=\pm\frac{\alpha\sqrt{N-m}}{\sqrt{m}} \tag{C.2}$$

and

$$\gamma=\mp\frac{\beta\sqrt{m}}{\sqrt{N-m}} \tag{C.3}$$

In conclusion, with each choice of (α,β) as a point on the ellipse, there correspond two points (γ,δ) (with oppositive sign) on the same ellipse, for which we obtain two orthogonal vectors $(\alpha\beta=-\gamma\delta)$. All other eigenvectors are orthogonal on x_1 and x_2. Thus, $x_k^T=\begin{bmatrix} v_k & w_k \end{bmatrix}^T$ obeys $x_k^T x_1=0$ and $x_k^T x_2=0$,

$$\begin{cases} \alpha v_k^T u+\beta w_k^T u=0 \\ \gamma v_k^T u+\delta w_k^T u=0 \end{cases}$$

or

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}\begin{bmatrix} v_k^T u \\ w_k^T u \end{bmatrix}=0$$

which only has the zero solution $v_k^T u=w_k^T u=0$ because $\det\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}=\frac{1}{\sqrt{(N-m)m}}>0$. Since all other eigenvectors x_k are orthogonal to u (with $\sum_{j=1}^N (x_k)_j=u^T x_k=0$), the sum of the elements in Z equals the sum of the elements in x_1 and x_2:

$$s_Z=(\alpha+\gamma)(N-m)+(\beta+\delta)m$$

Introducing the expression (C.3) for γ and (C.2) for δ into s_Z gives us

$$s_Z=\alpha N+(\alpha-\beta)\sqrt{m}\left(\sqrt{N-m}-\sqrt{m}\right)$$

From $1=\alpha^2(N-m)+\beta^2 m$, we eliminate $\alpha=\sqrt{\frac{1-\beta^2 m}{N-m}}$ and, after substitution, we have

$$s_Z=N\sqrt{\frac{1-\beta^2 m}{N-m}}+\left(\sqrt{\frac{1-\beta^2 m}{N-m}}-\beta\right)\sqrt{m}\left(\sqrt{N-m}-\sqrt{m}\right)$$

illustrating that, if $m=0$ and the graph is connected, then $s_Z=\sqrt{N}$. Moreover, s_Z is a function of the integer m and the real number β. For the case $1\le m<\frac{N}{2}$, it is convenient to denote $y=\beta^2 m\in(0,1)$ and write

$$s_Z(m,y)=N\sqrt{\frac{1-y}{N-m}}+\left(\sqrt{\frac{1-y}{N-m}}-\frac{\sqrt{y}}{\sqrt{m}}\right)\sqrt{m}\left(\sqrt{N-m}-\sqrt{m}\right)$$

For $y = 0$, we have $s_Z(m,0) = \sqrt{N-m}+\sqrt{m}$. Since $\left(\sqrt{N-m}+\sqrt{m}\right)^2 = N+2\sqrt{m}\sqrt{N-m} > N$, we find that $s_Z(m,0) > \sqrt{N}$. The other extremum $s_Z(m,1) = -\left(\sqrt{N-m}-\sqrt{m}\right)$is smaller than $s_Z(m,1) < 0 < \sqrt{N}$. Since y is a continuous real variable and $s_Z\left(m,y\right)$ is monotonously decreasing in y, there must exist, for each integer $m \in [1, \frac{N}{2})$, a $y^* \in (0,1)$ for which $s_Z\left(m,y^*\right) = \sqrt{N}$. In summary, we have demonstrated the following Theorem:

Theorem 14. *If the graph G is connected, then the number s_Z of elements in the orthogonal matrix Z of the Laplacian of the graph G equals $s_Z = \sqrt{N}$. The converse, "if $s_Z = \sqrt{N}$, then the graph G is connected" is not always true.*

D

CASCADING FAILURES IN INTERDEPENDENT NETWORKS

This appendix describes algorithms in Chapter 7. In addition, the derivative of the largest mutually connected component if all interlinks are alive, is presented.

D.1. ALGORITHMS: CASCADING AND COMPONENT

Algorithm 2 describes the function of cascading failures in interdependent networks. Lines 3 to 5 initialize a flag vector with flag = 1 if a node is not removed, otherwise flag = 0. Lines 6 to 9 remove the desired fraction $1-q$ of nodes and set flag = 0 for removed nodes. Due to the interconnection structure, the initial failures cause dependent nodes to fail executed by lines 13 to 26. As specified in line 18, a node u in G_1 is removed if it does not belong to the largest mutually connected component C_{G_1} or it loses all the dependent nodes. The same rule is applied for a node in G_2 as shown in line 23. Lines 18 and 23 correspond to the scenario of at least one interdependent node alive. The failure of a node u may introduce further failures and may invoke a cascading failure (line 11 is true). The cascading process is terminated if no more nodes fail and delNodes (in line 12) is not changed. Line 28 returns the resulting graph after removing all the failed nodes.

Algorithm 2 Function CASCADING(G, $1-q$, $\mathcal{N}_1$, $\mathcal{N}_2$)

1: **Input**: Graph G and fraction of removal $1-q$; Sets $\mathcal{N}_1$, $\mathcal{N}_2$ of nodes in G_1 and G_2, respectively
2: **Output**: endGraph: a graph after removing all the failed nodes from G
3: **for** each node $u \in G$ **do**
4: flag[u] $\leftarrow 1$
5: **end for**
6: **for** $i = 1$ **to** $\lceil (1-q)N \rceil$ **do**
7: $G \leftarrow G \backslash \{u_1, u_2, \cdots, u_i\}$ {u_i is a randomly chosen node from graph G_1}
8: flag[$u_1, u_2, \cdots, u_i$] $\leftarrow 0$;
9: **end for**
10: delNodes $\leftarrow 1$
11: **while** delNodes $\neq 0$ **do**
12: delNodes $\leftarrow 0$
13: **for** each node $u \in G$ **do**
14: LMCC $\leftarrow$ COMPONENT($G, \mathcal{N}_1, \mathcal{N}_2$)
15: $C_{G_1} \leftarrow \mathcal{N}_1 \cap$ LMCC
16: $C_{G_2} \leftarrow \mathcal{N}_2 \cap$ LMCC
17: $N[u] \leftarrow$ get neighbors of u
18: **if** $u \in \mathcal{N}_1$ and $\left(u \notin C_{G_1}$ or $N[u] \cap C_{G_2} = \varnothing\right)$ and flag[u]=1 **then**
19: endGraph $\leftarrow G \backslash \{u\}$
20: flag[u] $\leftarrow 0$
21: delNodes $\leftarrow 1$
22: $G \leftarrow$ endGraph
23: **else if** $u \in \mathcal{N}_2$ and $\left(u \notin C_{G_2}$ or $N[u] \cap C_{G_1} = \varnothing\right)$ and flag[u]=1 **then**
24: repeat lines 18-21
25: **end if**
26: **end for**
27: **end while**
28: **return** endGraph

Algorithm 3 extracts the largest mutually connected component from a given graph G. In line 3, we first obtain all the connected components C_i of G with sizes in descending order. Then, lines 4 to 9 return the first connected component that includes nodes both in G_1 and G_2.

Algorithm 3 Function COMPONENT($G, \mathscr{N}_1, \mathscr{N}_2$)

1: **Input**: Graph G; Sets $\mathscr{N}_1$, $\mathscr{N}_2$ of nodes in G_1 and G_2, respectively
2: **Output**: Largest mutually connected component LMCC
3: Get connected components C_1, C_2, …, C_N of G ordered as $|C_1| \leq \ldots \leq |C_N|$
4: **for** $i = 1$ **to** N **do**
5: **if** $C_i \cap \mathscr{N}_1 \neq \varnothing$ **and** $C_i \cap \mathscr{N}_2 \neq \varnothing$ **then**
6: LMCC $\leftarrow C_i$
7: **break**
8: **end if**
9: **end for**
10: **return** LMCC

D.2. DERIVATIVE OF THE LARGEST MUTUALLY CONNECTED COMPONENT IF ALL INTERLINKS ARE ALIVE

Theorem 15. *Consider an interdependent network consisting of two graphs G_1 and G_2. The interconnection topology between graphs G_1 and G_2 is the random geometric graph. Assume a node is alive when all of its interdependent nodes are alive. The fraction S_i ($i = 1,2$) of the largest mutually connected component as a function of $1-q$ removals is approximated by*

$$S_1 = q\left(1-\varphi_{C_{G_1}}(1)\right)\exp\left(-p_{ij}N\varphi_{C_{G_2}}(1)\right) \tag{D.1}$$

$$S_2 = \left(1-\varphi_{C_{G_2}}(1)\right)\exp\left(p_{ij}N\left(q-q\varphi_{C_{G_1}}(1)-1\right)\right) \tag{D.2}$$

where

$$\begin{cases} \varphi_{C_{G_1}}(1) = \varphi_{D_{G_1}}\left(1-q\exp\left(-p_{ij}Nu_B\right)(1-u_A)\right) \\ \varphi_{C_{G_2}}(1) = \varphi_{D_{G_2}}\left(1-\exp\left(p_{ij}N\left(q-qu_A-1\right)\right)(1-u_B)\right) \end{cases}$$

and

$$\begin{cases} u_A = \varphi_{(D_{l^+}-1)}\left(1 - q\exp\left(-p_{ij}Nu_B\right)(1-u_A)\right) \\ u_B = \varphi_{(D_{l^+}-1)}\left(1 - \exp\left(p_{ij}N\left(q - qu_A - 1\right)\right)(1-u_B)\right) \end{cases}$$

where p_{ij} is the probability that there is a link l_{ij} between node i in graph G_1 and node j in graph G_2. $1-\varphi_{C_{G_1}}(1)$ is the fraction of nodes belonging to the giant component in graph G_1 and $1-\varphi_{C_{G_2}}(1)$ in graph G_2.

Proof. For a node n in G_1 with k dependent nodes in G_2, the probability that all the dependent nodes are alive follows

$$\sum_{k=0}^{\infty} \Pr[D_B = k]\,(1-u_B)^k$$

which can be written as the generating function $\varphi_{D_B}(1-u_B)$ of D_B with parameter $1-u_B$. Assuming D_B follows a binomial distribution, it holds [33] that $\varphi_{D_B}(1-u_B) = \exp(-E[D_B]u_B)$ for a large interconnection matrix B. When B is the random geometric graph, the degree distribution of D_B follows a binomial distribution [169] with average degree $E[D_B] = p_{ij}N$. Therefore, the probability that all the dependent nodes in G_2 of a node n in G_1 are alive is $\exp\left(-p_{ij}Nu_B\right)$.

The self-consistent equation for u_A in interdependent network with one-to-many interconnection follows

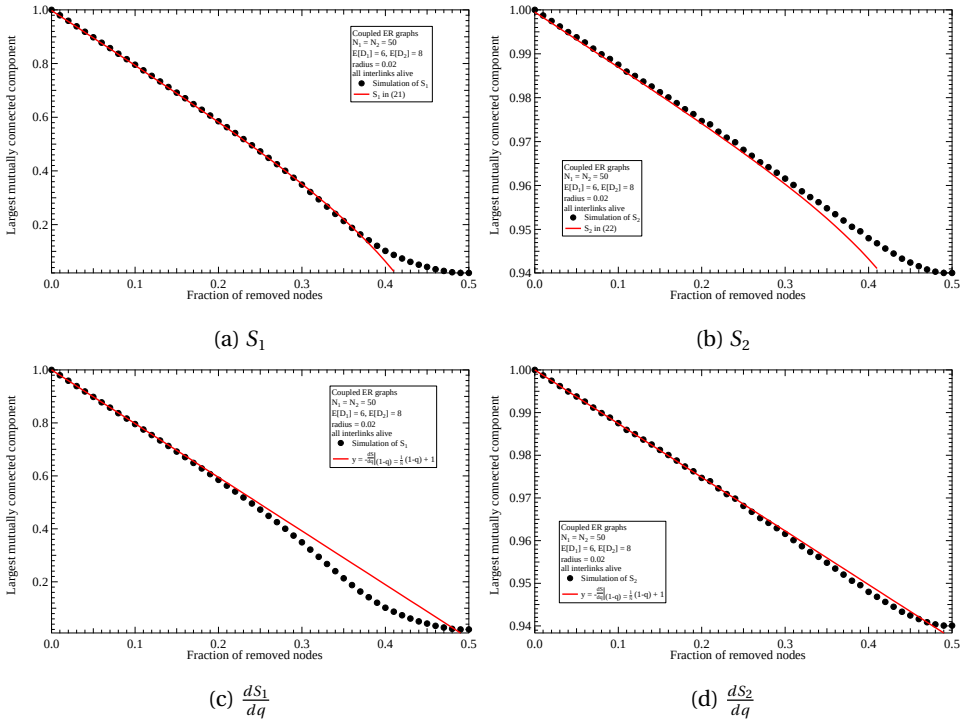
$$u_A = \varphi_{(D_{l^+}-1)}\left(1 - q\exp\left(-p_{ij}Nu_B\right)(1-u_A)\right)$$

where q is the probability for a node n to be occupied, and $\exp\left(-p_{ij}Nu_B\right)$ is the probability that all the interdependent nodes of a node n in G_1 belong to the giant component in graph G_2. Analogously,

$$u_B = \varphi_{(D_{l^+}-1)}\left(1 - \exp\left(p_{ij}N\left(q - qu_A - 1\right)\right)(1-u_B)\right)$$

Since we do not remove nodes from graph G_2 at the beginning, nodes in graph G_2 are occupied with probability 1. The probability $\exp\left(p_{ij}N\left(q - qu_A - 1\right)\right)$ represents that all the dependent nodes of a node in G_2 are occupied and belong to the giant component in G_1. □

For the scenario of all interdependent nodes alive, Figures D.1(a) and D.1(b) show the simulation results and S_1 and S_2 in (D.1) and (D.2) in coupled ER graphs with interconnection of random geometric graph with radius $r = 0.02$. Figures D.1(c,d)



(a) S_1 (b) S_2

(c) $\frac{dS_1}{dq}$ (d) $\frac{dS_2}{dq}$

Figure D.1: Largest mutually connected component as a function of the fraction of removed nodes in interdependent networks. The coupled graphs are Erdős-Rényi graphs $G_p(N)$ with $N = 50$ and the average degrees $E[D_1] = 6$ and $E[D_2] = 8$. The interconnection topology is the random geometric graph with $r = 0.02$. The results are averaged over 10^4 realizations of interdependent graphs.

show the simulation results and a straight line $y = -\frac{dS_i}{dq}\Big|_{1-q=\frac{1}{N}} \left(1-q\right) + 1$ $(i = 1,2)$, where the derivative $\frac{dS_i}{dq}$ $(i = 1,2)$ is numerically computed based on (D.1) and (D.2). In Figures D.1(c,d), the straight line with slope $-\frac{dS_i}{dq}$ $(i = 1,2)$ obtained from Theorem 15 shows a good approximation for the simulations for a small fraction of removals.

REFERENCES

[1] M. G. H. Bell, *A game theory approach to measuring the performance reliability of transport networks,* Transportation Research Part B: Methodological **34**, 533 (2000).

[2] S. Blumsack, L. B. Lave, and M. Ilić, *A quantitative analysis of the relationship between congestion and reliability in electric power networks,* The Energy Journal , 73 (2007).

[3] A. Leon-Garcia and I. Widjaja, *Communication networks* (McGraw-Hill, Inc., 2003).

[4] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, *Breakdown of the Internet under intentional attack,* Physical review letters **86**, 3682 (2001).

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

[6] R. Albert, *Scale-free networks in cell biology,* Journal of Cell Science **118**, 4947 (2005).

[7] Eaton, *Blackout and power outage tracker* (Eaton Corporation, 2015).

[8] U.S. Department of Energy and Natural Resources Canada, *Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations,* (April 2004).

[9] M. W. Horner and M. J. Widener, *The effects of transportation network failure on people's accessibility to hurricane disaster relief goods: a modeling approach and application to a florida case study,* Natural Hazards **59**, 1619 (2011).

[10] Y. Iida, *Basic concepts and future directions of road network reliability analysis,* Journal of advanced transportation **33**, 125 (1999).

[11] C.-L. Wu, *Improving airline network robustness and operational reliability by sequential optimisation algorithms,* Networks and Spatial Economics **6**, 235 (2006).

[12] A. K. Rose and M. M. Spiegel, *Cross-country causes and consequences of the 2008 crisis: early warning,* Japan and the World Economy **24**, 1 (2012).

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

[14] D. Y. Kenett, J. Gao, X. Huang, S. Shao, I. Vodenska, S. V. Buldyrev, G. Paul, H. E. Stanley, and S. Havlin, *Network of interdependent networks: overview of theory and applications,* in *Networks of Networks: The Last Frontier of Complexity* (Springer, 2014) pp. 3–36.

[15] M. Parandehgheibi and E. Modiano, *Robustness of interdependent networks: The case of communication networks and the power grid,* in *2013 IEEE Global Communications Conference (GLOBECOM)* (IEEE, 2013) pp. 2164–2169.

[16] P. Erdős and A. Rényi, *On random graphs I,* Publ. Math. Debrecen **6**, 290 (1959).

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

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

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

[20] S. R. Broadbent and J. M. Hammersley, *Percolation processes,* in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 53 (Cambridge University Press, 1957) pp. 629–641.

[21] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Network robustness and fragility: Percolation on random graphs,* Physical review letters **85**, 5468 (2000).

[22] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Resilience of the Internet to random breakdowns,* Physical review letters **85**, 4626 (2000).

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

[24] A. Vázquez and Y. Moreno, *Resilience to damage of graphs with degree correlations,* Physical Review E **67**, 015101 (2003).

[25] H. S. Wilf, *generatingfunctionology* (Elsevier, 2013).

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

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

[28] Y. Koç, M. Warnier, P. Van Mieghem, R. E. Kooij, and F. M. T. Brazier, *The impact of the topology on cascading failures in a power grid model,* Physica A: Statistical Mechanics and its Applications **402**, 169 (2014).

[29] X. Wang, Y. Koç, R. E. Kooij, and P. Van Mieghem, *A network approach for power grid robustness against cascading failures,* in *7th International Workshop on Reliable Networks Design and Modeling (RNDM)* (IEEE, 2015) pp. 208–214.

[30] F. Kuipers and P. Van Mieghem, *MAMCRA: a constrained-based multicast routing algorithm,* Computer Communications **25**, 802 (2002).

[31] W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics,* in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, Vol. 115 (The Royal Society, 1927) pp. 700–721.

[32] M. Barahona and L. M. Pecora, *Synchronization in small-world systems,* Physical review letters **89**, 054101 (2002).

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

[34] X. Wang and P. Van Mieghem, *Orthogonal eigenvector matrix of the laplacian,* in *11th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)* (2015) pp. 358–365.

[35] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Catastrophic cascade of failures in interdependent networks,* Nature **464**, 1025 (2010).

[36] H. Wang, Q. Li, G. D' Agostino, S. Havlin, H. E. Stanley, and P. Van Mieghem, *Effect of the interconnected network structure on the epidemic threshold,* Physical Review E **88**, 022801 (2013).

[37] W. Ellens, F. M. Spieksma, P. Van Mieghem, A. Jamakovic, and R. E. Kooij, *Effective graph resistance,* Lin. Algebra Appl. **435**, 2491 (2011).

[38] A. Asztalos, S. Sreenivasan, B. K. Szymanski, and G. Korniss, *Distributed flow optimization and cascading effects in weighted complex networks,* Eur. Phys. J. B **85**, 1 (2012).

[39] W. Abbas and M. Egerstedt, *Robust graph topologies for networked systems,* in *3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems* (2012) pp. 85–90.

[40] A. Ghosh, S. Boyd, and A. Saberi, *Minimizing effective resistance of a graph,* SIAM Rev. **50**, 37 (2008).

[41] P. Van Mieghem, X. Ge, P. Schumm, S. Trajanovski, and H. Wang, *Spectral graph analysis of modularity and assortativity,* Phys. Rev. E **82**, 056113 (2010).

[42] P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, UK, 2011).

[43] M. Fiedler, *Algebraic connectivity of graphs,* Czech. Math. J. **23**, 298 (1973).

[44] R. Grone, R. Merris, and V. S. Sunder, *The laplacian spectrum of a graph,* SIAM J. Matrix Anal. Appl. **11**, 218 (1990).

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

[46] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, *Characterization of complex networks: A survey of measurements,* Adv. Phys. **56**, 167 (2007).

[47] R. E. Kooij, A. Jamakovic, F. van Kesteren, T. de Koning, I. Theisler, and P. Veldhoven, *The Dutch soccer team as a social network,* Connections **29** (2009).

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

[49] A. Jamakovic and S. Uhlig, *On the relationships between topological measures in real-world networks,* Networks and Heterogeneous Media **3**, 345 (2008).

[50] P. Gupta and P. R. Kumar, *The capacity of wireless networks,* IEEE Trans. Inf. Theory **46**, 388 (2000).

[51] R. K. Ganti and M. Haenggi, *Interference and outage in clustered wireless ad hoc networks,* IEEE Trans. Inf. Theory **55**, 4067 (2009).

[52] H. Wang and P. Van Mieghem, *Algebraic connectivity optimization via link addition,* in *Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Sytems* (ICST, 2008) p. 22.

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

[54] J. G. F. Francis, *The QR transformation - part 2,* Comput. J. **4**, 332 (1962).

[55] J. Dean and S. Ghemawat, *Mapreduce: simplified data processing on large clusters,* Comm. ACM **51**, 107 (2008).

[56] T. Watanabe and N. Masuda, *Enhancing the spectral gap of networks by node removal,* Phys. Rev. E **82**, 046102 (2010).

[57] A. Milanese, J. Sun, and T. Nishikawa, *Approximating spectral impact of structural perturbations in large networks,* Phys. Rev. E **81**, 046112 (2010).

[58] 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**, 319 (2012).

[59] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, *Attack vulnerability of complex networks,* Phys. Rev. E **65**, 056109 (2002).

[60] F. Dörfler and F. Bullo, *Synchronization of power networks: Network reduction and effective resistance,* in *IFAC Workshop on Distributed Estimation and Control in Networked Systems* (2010) pp. 197–202.

[61] D. Liben-Nowell and J. Kleinberg, *The link-prediction problem for social networks,* J. Am. Soc. Inform. Sci. Tech. **58**, 1019 (2007).

[62] L. Lü and T. Zhou, *Link prediction in complex networks: A survey,* Phys. Stat. Mech. Appl. **390**, 1150 (2011).

[63] M. E. J. Newman, *Clustering and preferential attachment in growing networks,* Phys. Rev. E **64**, 025102 (2001).

[64] J. W. Wang and L. L. Rong, *Robustness of the western United States power grid under edge attack strategies due to cascading failures,* Safety science **49**, 807 (2011).

[65] L. Chang and Z. Wu, *Performance and reliability of electrical power grids under cascading failures,* International Journal of Electrical Power & Energy Systems **33**, 1410 (2011).

[66] Y. Koç, M. Warnier, R. E. Kooij, and F. M. T. Brazier, *An entropy-based metric to quantify the robustness of power grids against cascading failures,* Safety science **59**, 126 (2013).

[67] P. Hines, K. Balasubramaniam, and E. C. Sanchez, *Cascading failures in power grids,* Potentials, IEEE **28**, 24 (2009).

[68] I. Dobson, *Cascading network failure in power grid blackouts,* in *Encyclopedia of Systems and Control* (Springer London, 2014) pp. 1–5.

[69] G. Zhang, Z. Li, B. Zhang, and W. A. Halang, *Understanding the cascading failures in Indian power grids with complex networks theory,* Physica A: Statistical Mechanics and its Applications **392**, 3273 (2013).

[70] B. Karrer, M. E. J. Newman, and L. Zdeborová, *Percolation on sparse networks,* Phys. Rev. Lett. **113**, 208702 (2014).

[71] S. Trajanovski, J. Martín-Hernández, W. Winterbach, and P. Van Mieghem, *Robustness envelopes of networks,* Journal of Complex Networks **1**, 44 (2013).

[72] W. Ellens, F. M. Spieksma, P. Van Mieghem, A. Jamakovic, and R. E. Kooij, *Effective graph resistance,* Lin. Algebra Appl. **435**, 2491 (2011).

[73] D. Braess, A. Nagurney, and T. Wakolbinger, *On a paradox of traffic planning,* Transportation science **39**, 446 (2005).

[74] J. J. Grainger and W. D. Stevenson, *Power system analysis,* McGraw-Hill series in electrical and computer engineering (1994).

[75] J. G. Restrepo, E. Ott, and B. R. Hunt, *Approximating the largest eigenvalue of network adjacency matrices,* Phys. Rev. E **76**, 056119 (2007).

[76] M. E. J. Newman, *Community detection and graph partitioning,* EPL(Europhysics Letters) **103**, 28003 (2013).

[77] X. Wang, E. Pournaras, R. E. Kooij, and P. Van Mieghem, *Improving robustness of complex networks via the effective graph resistance,* The European Physical Journal B **87**, 1 (2014).

[78] V. Latora and M. Marchiori, *Vulnerability and protection of infrastructure networks,* Physical Review E **71**, 015103 (2005).

[79] Y. Koç, T. Verma, N. A. M. Araujo, and M. Warnier, *MATCASC: A tool to analyse cascading line outages in power grids,* in *Intelligent Energy Systems (IWIES), IEEE International Workshop on* (IEEE, 2013) pp. 143–148.

[80] B. Calvert and G. Keady, *Braess's paradox and power-law nonlinearities in networks,* The Journal of the Australian Mathematical Society, Series B. Applied Mathematics **35**, 1 (1993).

[81] S. Wernicke and F. Rasche, *FANMOD: a tool for fast network motif detection,* Bioinformatics **22**, 1152 (2006).

[82] United Nations, *World Urbanization Prospects: The 2014 Revision* (United Nations, New York, 2014).

[83] V. R. Vuchic, *Urban transit systems and technology* (John Wiley & Sons, 2007).

[84] S. Derrible, *Urban infrastructure is not a tree: Integrating and decentralizing urban infrastructure systems,* Environment and Planning B: Planning and Design (2016).

[85] S. Derrible, *Complexity in future cities: the rise of networked infrastructure,* International Journal of Urban Sciences , 1 (2016).

[86] S. Derrible, S. Saneinejad, L. Sugar, and C. Kennedy, *Macroscopic model of greenhouse gas emissions for municipalities,* Transportation Research Record: Journal of the Transportation Research Board **2191**, 174 (2010).

[87] C. D. Cottrill and S. Derrible, *Leveraging Big Data for the Development of Transport Sustainability Indicators,* Journal of Urban Technology **22**, 45 (2015).

[88] B. M. Ayyub, *Systems resilience for multihazard environments: Definition, metrics, and valuation for decision making,* Risk Analysis **34**, 340 (2014).

[89] W. Cai, S. Borlace, M. Lengaigne, P. Van Rensch, M. Collins, G. Vecchi, A. Timmermann, A. Santoso, M. J. McPhaden, L. Wu, *et al.*, *Increasing frequency of extreme el niño events due to greenhouse warming,* Nature Climate Change (2014).

[90] A. Kermanshah and S. Derrible, *A geographical and multi-criteria vulnerability assessment of transportation networks against extreme earthquakes,* Reliability Engineering & System Safety **153**, 39 (2016).

[91] D. D. Woods, *Four concepts for resilience and the implications for the future of resilience engineering,* Special Issue on Resilience Engineering **141**, 5 (2015).

[92] H. S. Levinson, *The reliability of transit service: An historical perspective,* Journal of Urban Technology **12**, 99 (2005).

[93] L. M. Kieu, A. Bhaskar, and E. Chung, *Public transport travel-time variability definitions and monitoring,* Journal of Transportation Engineering **141** (2015).

[94] E. Mazloumi, G. Currie, and G. Rose, *Using GPS data to gain insight into public transport travel time variability,* Journal of Transportation Engineering **136**, 623 (2010).

[95] V. Benezech and N. Coulombel, *The value of service reliability,* Transportation Research Part B: Methodological **58**, 1 (2013).

[96] B. Yao, P. Hu, X. Lu, J. Gao, and M. Zhang, *Transit network design based on travel time reliability,* Transportation Research Part C: Emerging Technologies **43**, 233 (2014).

[97] K. An and H. K. Lo, *Service reliability based transit network design with stochastic demand,* in *Transportation Research Board 93rd Annual Meeting,* 14-1971 (2014).

[98] M. E. J. Newman, *Networks: an introduction* (Oxford University Press, 2010).

[99] D. J. Watts, *A simple model of global cascades on random networks,* Proceedings of the National Academy of Sciences **99**, 5766 (2002).

[100] P. Crucitti, V. Latora, and M. Marchiori, *Model for cascading failures in complex networks,* Physical Review E **69**, 045104 (2004).

[101] R. Kinney, P. Crucitti, R. Albert, and V. Latora, *Modeling cascading failures in the North American power grid,* The European Physical Journal B-Condensed Matter and Complex Systems **46**, 101 (2005).

[102] S. Derrible and C. Kennedy, *A network analysis of subway systems in the world using updated graph theory,* Transportation Research Record **2112**, 17 (2009).

[103] S. Derrible and C. Kennedy, *Characterizing metro networks: State, Form, and Structure,* Transportation **37**, 275 (2010).

[104] B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, *Resilience of public transport networks against attacks,* European Physical Journal B **71**, 125 (2009).

[105] S. Derrible and C. Kennedy, *The complexity and robustness of metro networks,* Physica A: Statistical Mechanics and its Applications **389**, 3678 (2010).

[106] C. von Ferber, B. Berche, T. Holovatch, and Y. Holovatch, *A tale of two cities: Vulnerabilities of the London and Paris transit networks,* Journal of Transportation Security **5**, 199 (2012).

[107] E. Rodríguez-Núñez and J. C. García-Palomares, *Measuring the vulnerability of public transport networks,* Journal of Transport Geography **35**, 50 (2014).

[108] H. Kim, C. Kim, and Y. Chun, *Network Reliability and Resilience of Rapid Transit Systems,* The Professional Geographer , 1 (2015).

[109] X. Wang, Y. Koç, S. Derrible, S. N. Ahmad, and R. E. Kooij, *Quantifying the robustness of metro networks,* in *6th International Symposium on Transportation Network Reliability* (Nara, Japan, 2015).

[110] O. Cats and E. Jenelius, *Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information,* Networks and Spatial Economics **14**, 435 (2014).

[111] O. Cats and E. Jenelius, *Planning for the unexpected: The value of reserve capacity for public transport network robustness,* Transportation Research Part A: Policy and Practice **81**, 47 (2015).

[112] S. Derrible, *Network centrality of metro systems,* PloS ONE **7**, e40575 (2012).

[113] C. J. Colbourn, *The combinatorics of network reliability*, Vol. 200 (Oxford University Press New York, 1987).

[114] J. Wu, M. Barahona, Y. J. Tan, and H. Z. Deng, *Spectral measure of structural robustness in complex networks,* Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on **41**, 1244 (2011).

[115] 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 **2011**, P11018 (2011).

[116] A. Cardillo, S. Scellato, V. Latora, and S. Porta, *Structural properties of planar graphs of urban street patterns,* Physical Review E **73**, 066107 (2006).

[117] R. Cohen and S. Havlin, *Complex networks: structure, robustness and function* (Cambridge University Press, 2010).

[118] M. D. Yap, *Robust public transport from a passenger perspective: A study to evaluate and improve the robustness of multi-level public transport networks*, Master's thesis, Delft University of Technology (2014).

[119] M. L. Rebaiaia and D. Ait-Kadi, *Network Reliability Evaluation and Optimization: Methods, Algorithms and Software Tools* (Interuniversity research center on enterprise networks, logistics and transportation, 2013).

[120] M. O. Ball, C. J. Colbourn, and J. S. Provan, *Network reliability,* Handbooks in operations research and management science **7**, 673 (1995).

[121] S. Tahmasseby, *Reliability in urban public transport network assessment and design* (TU Delft, Delft University of Technology, 2009).

[122] J. Carlier and C. Lucet, *A decomposition algorithm for network reliability evaluation,* Discrete Applied Mathematics **65**, 141 (1996).

[123] R. K. Kincaid, N. Alexandrov, and M. J. Holroyd, *An investigation of synchrony in transport networks,* Complexity **14**, 34 (2009).

[124] P. Wei, G. Spiers, and D. Sun, *Algebraic connectivity maximization for air transportation networks,* Intelligent Transportation Systems, IEEE Transactions on **15**, 685 (2014).

[125] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, *Epidemic outbreaks in complex heterogeneous networks,* The European Physical Journal B-Condensed Matter and Complex Systems **26**, 521 (2002).

[126] M. E. J. Newman, I. Jensen, and R. M. Ziff, *Percolation and epidemics in a two-dimensional small world,* Physical Review E **65**, 021904 (2002).

[127] N. Taleb, *Antifragile: Things That Gain from Disorder* (Random House Publishing Group, New York, NY, 2012).

[128] P. Van Mieghem, C. Doerr, H. Wang, J. Martin Hernandez, D. Hutchison, M. Karaliopoulos, and R. E. Kooij, *A framework for computing topological network robustness,* in *Technical Report 20101218, Delft University of Technology* (2010).

[129] R. Cervero, *The transit metropolis: a global inquiry* (Island press, 1998).

[130] R. Albert, I. Albert, and G. L. Nakarado, *Structural vulnerability of the North American power grid,* Phys. Rev. E **69**, 025103 (2004).

[131] M. J. Krawczyk, L. Muchnik, A. Manka-Krason, and K. Kulakowski, *Line graphs as social networks,* Physica A **390**, 2611 (2011).

[132] J. B. Pereira-Leal, A. J. Enright, and C. A. Ouzounis, *Detection of functional modules from protein interaction networks,* PROTEINS: Structure, Function, and Bioinformatics **54**, 49 (2004).

[133] R. Diestel, *Graph theory,* Grad. Texts in Math (2005).

[134] T. S. Evans and R. Lambiotte, *Line graphs, link partitions, and overlapping communities,* Physical Review E **80**, 016105 (2009).

[135] J. C. Wierman, D. P. Naor, and J. Smalletz, *Incorporating variability into an approximation formula for bond percolation thresholds of planar periodic lattices,* Physical Review E **75**, 011114 (2007).

[136] H. Whitney, *Congruent graphs and the connectivity of graphs,* American Journal of Mathematics **54**, 150 (1932).

[137] J. Krausz, *Démonstration nouvelle d'un théorème de Whitney sur les réseaux,* Mat. Fiz. Lapok **50**, 75 (1943).

[138] A. C. M. van Rooij and H. S. Wilf, *The interchange graph of a finite graph,* Acta Mathematica Academiae Scientiarum Hungaricae **16**, 263 (1965).

[139] F. Harary, *Graph Theory* (Addison-Wesley, 1969).

[140] D. Liu, S. Trajanovski, and P. Van Mieghem, *Random line graphs and a linear law for assortativity,* Physical Review E **87**, 012816 (2013).

[141] N. D. Roussopoulos, *A $max\{m,n\}$ algorithm for detecting the graph H from its line graph G,* Information Processing Letters **2**, 108 (1973).

[142] P. G. H. Lehot, *An optimal algorithm to detect a line graph and output its root graph,* Journal of the ACM **21**, 569 (1974).

[143] D. Liu, S. Trajanovski, and P. Van Mieghem, *ILIGRA: an efficient inverse line graph algorithm,* Journal of Mathematical Modelling and Algorithms in Operations Research **14**, 13 (2014).

[144] E. R. van Dam and W. H. Haemers, *Which graphs are determined by their spectrum?* Linear Algebra and its applications **373**, 241 (2003).

[145] 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 **76**, 643 (2010).

[146] M. E. J. Newman, *Assortative mixing in networks,* Physical review letters **89**, 208701 (2002).

[147] R. Noldus and P. Van Mieghem, *Assortativity in complex networks,* Journal of Complex Networks (2015).

[148] M. Youssef, Y. Khorramzadeh, and S. Eubank, *Network reliability: the effect of local network structure on diffusive processes,* Physical Review E **88**, 052810 (2013).

[149] V. Colizza, R. Pastor-Satorras, and A. Vespignani, *Reaction–diffusion processes and metapopulation models in heterogeneous networks,* Nature Physics **3**, 276 (2007).

[150] V. Batagelj and A. Mrvar, *Pajek datasets,* (2006).

[151] E. R. van Dam and R. E. Kooij, *The minimal spectral radius of graphs with a given diameter,* Linear Algebra and its Applications **423**, 408 (2007).

[152] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Complex networks: Structure and dynamics,* Physics reports **424**, 175 (2006).

[153] J. Martín-Hernández, *Measuring Robustness of Complex Networks* (TU Delft, Delft University of Technology, 2013).

[154] E. Estrada, *The structure of complex networks: theory and applications* (OUP Oxford, 2011).

[155] D. Stevanović, *Spectral Radius of Graphs* (Academic Press, 2014).

[156] A. Jamakovic and P. Van Mieghem, *On the robustness of complex networks by using the algebraic connectivity* (Springer, 2008).

[157] D. M. Cvetković, P. Rowlinson, and S. Simic, *Eigenspaces of graphs*, 66 (Cambridge University Press, 1997).

[158] D. M. Cvetković, P. Rowlinson, and S. Simić, *An Introduction to the Theory of Graph Spectra* (Cambridge University Press, 2010).

[159] A. Pothen, H. D. Simon, and K.-P. Liou, *Partitioning sparse matrices with eigenvectors of graphs,* SIAM journal on matrix analysis and applications **11**, 430 (1990).

[160] B. Bollobás, *Random graphs* (Springer, 1998).

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

[162] L. Glasser, K. T. Kohl, C. Koutschan, V. H. Moll, and A. Straub, *The integrals in Gradshteyn and Ryzhik. Part 22: Bessel-K functions,* Scientia. Series A. Mathematical Sciences. New Series **22**, 129 (2012).

[163] M. Abramowitz, I. A. Stegun, *et al.*, *Handbook of mathematical functions,* Applied mathematics series **55**, 62 (1966).

[164] V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. Porcellinis, and R. Setola, *Modelling interdependent infrastructures using interacting dynamical models,* International Journal of Critical Infrastructures **4**, 63 (2008).

[165] P. Van Mieghem, *Interconnectivity structure of a general interdependent network,* Physical Review E **93**, 042305 (2016).

[166] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, *The structure and dynamics of multilayer networks,* Physics Reports **544**, 1 (2014).

[167] F. Radicchi and A. Arenas, *Abrupt transition in the structural formation of interconnected networks,* Nature Physics **9**, 717 (2013).

[168] F. D. Sahneh, C. Scoglio, and P. Van Mieghem, *Exact coupling threshold for structural transition reveals diversified behaviors in interconnected networks,* Physical Review E **92**, 040801 (2015).

[169] J. Dall and M. Christensen, *Random geometric graphs,* Physical Review E **66**, 016121 (2002).

[170] G. T. Toussaint, *The relative neighbourhood graph of a finite planar set,* Pattern recognition **12**, 261 (1980).

[171] P. Van Mieghem, *Paths in the simple random graph and the waxman graph,* Probability in the Engineering and Informational Sciences **15**, 535 (2001).

[172] K. Kratky, *The area of intersection of N equal circular disks,* Journal of Physics A: Mathematical and General **11**, 1017 (1978).

[173] L. Devroye, *The expected size of some graphs in computational geometry,* Computers & mathematics with applications **15**, 53 (1988).

[174] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, *Networks formed from interdependent networks,* Nature physics **8**, 40 (2012).

[175] R. Parshani, S. V. Buldyrev, and S. Havlin, *Interdependent networks: reducing the coupling strength leads to a change from a first to second order percolation transition,* Physical review letters **105**, 048701 (2010).

[176] P. Crucitti, V. Latora, and M. Marchiori, *A topological analysis of the Italian electric power grid,* Physica A: Statistical mechanics and its applications **338**, 92 (2004).

[177] X.-Y. Li, P.-J. Wan, and O. Frieder, *Coverage in wireless ad hoc sensor networks,* IEEE Transactions on Computers **52**, 753 (2003).

[178] P. Van Mieghem, *Interconnectivity structure of a general interdependent network,* Physical Review E **93**, 042305 (2016).

[179] R. Little, *A socio-technical systems approach to understanding and enhancing the reliability of interdependent infrastructure systems,* International Journal of Emergency Management **2**, 98 (2004).

[180] A. Vespignani, *Complex networks: The fragility of interdependency,* Nature **464**, 984 (2010).

[181] S. Bonaccorsi, S. Ottaviano, F. De Pellegrini, A. Socievole, and P. Van Mieghem, *Epidemic outbreaks in two-scale community networks,* Physical Review E **90**, 012810 (2014).

[182] F. Radicchi and A. Arenas, *Abrupt transition in the structural formation of interconnected networks,* Nature Physics **9**, 717 (2013).

[183] J. Martín-Hernández, H. Wang, P. Van Mieghem, and G. D' Agostino, *Algebraic connectivity of interdependent networks,* Physica A: Statistical Mechanics and its Applications **404**, 92 (2014).

[184] F. D. Sahneh, C. Scoglio, and P. Van Mieghem, *Exact coupling threshold for structural transition reveals diversified behaviors in interconnected networks,* Physical Review E **92**, 040801 (2015).

[185] D. Ganesan, R. Cristescu, and B. Beferull-Lozano, *Power-efficient sensor placement and transmission structure for data gathering under distortion constraints,* ACM Transactions on Sensor Networks **2**, 155 (2006).

[186] V. K. Sood, D. Fischer, J. Eklund, and T. Brown, *Developing a communication infrastructure for the smart grid,* in *Electrical Power & Energy Conference (EPEC)* (IEEE, 2009).

[187] P. Tewarie, A. Hillebrand, B. W. van Dijk, C. J. Stam, G. C. O'Neill, P. Van Mieghem, J. M. Meier, M. W. Woolrich, P. G. Morris, and M. J. Brookes, *Integrating cross-frequency and within band functional networks in resting-state MEG: a multi-layer network approach,* NeuroImage **to appear** (2016).

[188] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, *Identifying, understanding, and analyzing critical infrastructure interdependencies,* Control Systems, IEEE **21**, 11 (2001).

[189] M. Fiedler, *Algebraic connectivity of graphs,* Czechoslovak mathematical journal **23**, 298 (1973).

[190] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardeñes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, *Diffusion dynamics on multiplex networks,* Physical review letters **110**, 028701 (2013).

[191] C. W. Wu, *Synchronization in arrays of coupled nonlinear systems with delay and nonreciprocal time-varying coupling,* IEEE Transactions on Circuits and Systems II: Express Briefs **52**, 282 (2005).

[192] I. Sciriha, *On the construction of graphs of nullity one,* Discrete Mathematics **181**, 193 (1998).

[193] F. Harary, *Graph theory,* Addision wesley, New Delhi (1969).

[194] F. M. Atay and T. Bıyıkoğlu, *Graph operations and synchronization of complex networks,* Physical Review E **72**, 016217 (2005).

[195] R. Merris, *Laplacian graph eigenvectors,* Linear algebra and its applications **278**, 221 (1998).

[196] P. J. Davis, *Circulant matrices* (American Mathematical Soc., 2012).

[197] P. J. Davis and P. Rabinowitz, *Methods of numerical integration* (Courier Corporation, 2007).

[198] H. Wang, W. Winterbach, and P. Van Mieghem, *Assortativity of complementary graphs,* The European Physical Journal B **83**, 203 (2011).

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my promoters and scientific fathers, Prof. Piet Van Mieghem and Prof. Robert E. Kooij who together shape me to be a scientific researcher and challenge my critical thinking.

I would like to share two stories between Piet and me. At the beginning of my PhD, he wrote down in my notebook Gauss' adage "Pauca, sed matura", which means "Few, but ripe". I always bear it in my mind and will keep the notebook as a reminder. Another story happened in our way to Aula for lunch. Piet seriously told me: use all your spare time to learn English. I end up with watching an American TV series for so many times that I can clearly memorize every single plot.

Special thanks to Rob. Thank you for taking me over from my previous daily supervisor after a half year of my Ph.D. and for establishing a good starting point for my Ph.D. You always tell me "play with it" which is the key to make my exploration deeper and more enjoyable. Thanks for encouraging me to conferences to give presentation and meet people. Thanks for your criticize on my presentation which I learned a lot.

I would like to express my gratitude to the committee members of my thesis defence for their time and effort spent on this thesis. Special thanks go to Dr. Evangelos Pournaras for his guidance and our nice conversations. My special thanks go to Dr. Edgar van Boven for translating the summary and the proposition into Dutch. I would like to warmly thank my office mates: Cong, Ruud, Marcus, Hale and Zhidong, for providing me such a happy working environment. Thanks to all the NASers: Nico, Fernando, Huijuan, Remco, Marloes, Wendy, Norbert, Niels, Farabi, Chuan, Rogier, Jil, Aleksandar, Qiang, Yakup, Song, Stojan, Wynand, Javier, Dajie, Annalisa. Spectial thanks go to NAS secretary Rani Ramnares for her help on defence preparations.

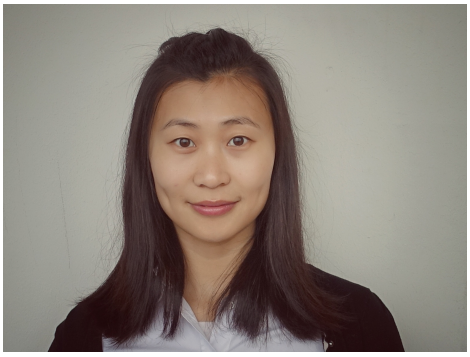
I want to thank the people who make my stay in Delft so enjoyable: Bo Qu, Haohua Zong, Jia Xu, Jialin Han, Jinjun Li, Lei Shi, Qingqing Ye, Wenjie Pei, Xiaosi

Xu, Yang zhou, Yazhou Yang, Yu Xin, Zi Wang and my friends from Latin group and Zumba group. To my two Chinese girls Huarong Zheng and Xiaoyan Wei, I had such a great time with you girls.

Finally, my warmest thanks to my parents and my older brother, Xinxin Wang, for their endless love and support. To my brother, I am always proud of our names: Xinxin Xiangrong which stands for flourishing.

Xiangrong Wang
Delft, December 2016

CURRICULUM VITÆ



Xiangrong Wang (王向荣) was born in Shandong (China) on 29 September, 1988 (though officially 19 August, 1986). She graduated from Yantai University with a Bachelor's degree in Telecommunication Engineering on July 2010. She obtained her Master's degree from Beijing Jiaotong University in Image Processing on July 2012. Supported by China Scholarship Council (CSC), she continued with her Ph.D. study. Since October 2012, she became a Ph.D. researcher in the Networks Architectures and Services (NAS) group, faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, under the supervision of Prof. Piet Van Mieghem and Prof. Robert E. Kooij.

Her research focuses are robustness of complex networks and fundamentals of graph theory. The research results have been published, some of them have been presented, in conferences and international journals. She assisted Prof. P. Van Mieghem with teaching the introduction of a master's course, Performance Analysis, and with supervising projects for the course of Networking. She has reviewed several Journal and conference papers.

LIST OF PUBLICATIONS

P.9 **X. Wang**, Ling Feng and Robert E. Kooij, 2016, *"Inconsistencies among spectral robustness metrics"*, Applicable analysis and discrete mathematics, submitted.

P.8 **X. Wang**, R. E. Kooij and P. Van Mieghem, 2016, *"Structural transition in interdependent networks with regular interconnections"*, Journal of Statistical Mechanics: Theory and Experiment (JSTAT), submitted.

P.7 **X. Wang**, R. E. Kooij and P. Van Mieghem, 2016, *"Modelling region-based interconnection for interdependent networks"*, Physical Review E 94(4), 042315.

P.6 **X. Wang**, Yakup Koç, Sybil Derrible, Sk Nasir Ahmad, Willem J.A. Pino and Robert E. Kooij, 2016, *"Multi-Criteria Robustness Analysis of Metro Networks"*, Physica A, submitted.

P.5 **X. Wang**, S. Trajanovski, R. E. Kooij and P. Van Mieghem, 2015, *"Degree distribution and assortativity in line graphs of complex networks"*, Physica A, vol. 445, pp. 343-356.

P.4 **X. Wang** and P. Van Mieghem, 2015, *"Orthogonal Eigenvector Matrix of the Laplacian"*, 11th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), November 23-27, Bangkok, Thailand.

P.3 **X. Wang**, Y. Koc, R. E. Kooij and P. Van Mieghem, 2015, *"A network approach for power grid robustness against cascading failures"* , RNDM 2015 - 7th International Workshop on Reliable Networks Design and Modeling, October 5-7, Munich, Germany.

P.2 **X. Wang**, Y. Koc, S. Derrible, S. N. Ahmad and R. E. Kooij, 2015, *"Quantifying the robustness of metro networks"* , INSTR 2015 - 6th International Symposium on Transportation Network Reliability, August 2-3, Nara, Japan.

P.1 **X. Wang**, E. Pournaras, R. E. Kooij and P. Van Mieghem, 2014, *" Improving robustness of complex networks via the effective graph resistance "*, The European Physical Journal B, 87 (9), 1-12.

CONFERENCES WITH ABSTRACT-ONLY

5. X. Wang, R. E. Kooij and P. Van Mieghem, 2016, *" Location based interconnection design for interdependent networks"* , Conference on Complex Systems, September 19-22, Amsterdam, The Netherlands.

4. X. Wang, R. E. Kooij and P. Van Mieghem, 2016, "*The algebraic connectivity of an interdependent network*", Spectra of graphs and applications, May 18–20, Serbian Academy of Sciences and Arts, Belgrade, Serbia.

3. R. E. Kooij, X. Wang and P. Van Mieghem, 2016, "*Inconsistencies among spectral robustness metrics*", Spectra of graphs and applications, May 18–20, Serbian Academy of Sciences and Arts, Belgrade, Serbia.

2. P. Van Mieghem, X. Wang and R. E. Kooij, 2016, "*Eigenvector matrix of the Laplacian*", Spectra of graphs and applications, May 18–20, Serbian Academy of Sciences and Arts, Belgrade, Serbia.

1. X. Wang, R. E. Kooij and P. Van Mieghem, 2015, "*Improving the robustness of complex networks via the effective graph resistance*", First Delft – Girona Workshop on Robustness of Networks, June 17-18, Girona, Spain.

RELATIONS TO THIS THESIS

This thesis consists of published articles of the author. Table D.1 presents the relation between the list of publications from P.1 to P.9 and the chapters of this thesis.

Table D.1: Relation between the list of publications and the chapters of this thesis.

Publication	Intro	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch. 6	Ch. 7	Ch. 8	Conclusion
P.1		●							
P.2				●					
P.3			●						
P.4						●			
P.5					●				
P.6				●					
P.7							●		
P.8								●	
P.9		●	●	●					