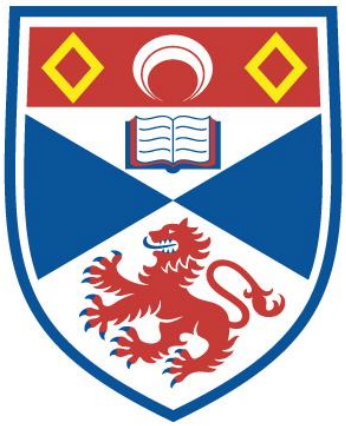
COUPLED COMPLEX NETWORKS: STRUCTURE, ADAPTATION, AND PROCESS

Saray Shai

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2014

Coupled complex networks: structure, adaptation, and processes

Saray Shai

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy

at the University of St Andrews

May 2014

Abstract

In the last 15 years, network science has established itself as a leading scientific tool for the study of complex systems, describing how components in a system interact with one another. Understanding the structure and dynamics of these networks of interactions is the key to understanding the global behaviour of the systems they represent, with a wide range of applications to fundamental societal problems; from designing stable and resilient infrastructures which are critical to our sustainability, to identifying topological patterns in interactome networks that are associated with breast cancer.

Most studies so far have focused on isolated single networks that do not interact with or depend upon other networks, while in reality networks rarely live in isolation and are often just one component in a much larger complex multilevel network. Together with the increased availability of richer, bigger and multi-relational datasets, the analysis of coupled networks has been recently attracting many researchers, and has exposed a multitude of new features and phenomena that were not observed for isolated networks.

In this thesis, we present analytical, numerical and empirical studies of coupled complex networks, aiming to understand the implications of coupling to the functionality and behaviour of complex systems.

First, we present a theoretical framework for studying the robustness of modular or interconnected networks, providing the critical concentration of interconnections between modules, above which the internal structure of each module is inseparable from the system as a whole. Second, we present another theoretical framework to study epidemic spreading in interconnected adaptive networks, discovering a new stationary state that only emerges in the case of weakly coupled networks, where the epidemic localise in the coupled nodes. In order to obtain the exact quantitative behavior of the new state from the analytical model, one must account for the actual second-order moments of the system, even for homogeneous networks, where in single networks it is usually sufficient to treat such higher-order terms by a uniform approximation. Thirdly, we present a numerical study on the effect of correlated coupling on spreading dynamics in the presence of resource constraints, finding that positive correlation between coupled nodes can impede flow process through contention, and thus constitute a less spreading-efficient structure than negatively correlated networks. Finally, we complete the thesis with a large-scale empirical study of interacting transportation networks in the entire metropolitan areas of both London and New York. We find that coupling can strongly affect the structure, and consequently the behaviour, of such multilayer transportation systems.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Simon Dobson for showing me how to enjoy research. This thesis would not have been possible without him and without his great advice, support and encouragement he has given me over the last four years.

I am also thankful for the further advice and inspiration I have received from my second supervisors Prof. Ian Gent and Prof. Jane Hillston.

I would like to thank my committee members Prof. Stephen Linton and Prof. David Arrowsmith for a stimulating discussion during my viva, and for their useful suggestions.

I also thank my friends and colleagues in the School of Computer Science, who are always there for me, supporting and inspiring my research, equally happy to have a stimulating discussion or an early afternoon pint.

Last but not least, I thank my family for their endless love and encouragement and for always believing in me.

Publications

Some of the work presented in this thesis has been previously published:

I Saray Shai and Simon Dobson *Effect of resource constraints on intersimilar coupled networks*. Phys Rev E 86(6): 066120 (2012)

II Saray Shai and Simon Dobson *Coupled adaptive complex networks*. Phys Rev E 87(4): 042812 (2013)

III Saray Shai, Dror Y. Kenett, Yoed N. Kenett, Miriam Faust, Simon Dobson and Shlomo Havlin *Resilience of modular complex networks*. Manuscript submitted for publication. arXiv:1404.4748 (2014)

IV Emanuele Strano, Saray Shai, Simon Dobson and Marc Barthélemy *Efficiency and centrality of multiplex spatial networks in large urban areas*. Manuscript submitted for publication (2014)

Declaration

Candidate's Declarations

I, Saray Shai, hereby certify that this thesis, which is approximately 31 113 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student and as a candidate for the degree of Doctor of Philosophy in September 2010; the higher study for which this is a record was carried out in the University of St Andrews between 2010 and 2014.

Date:

Signature of candidate:

Supervisor's Declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date:

Signature of supervisor:

Permission for Electronic Publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

> *Access to printed copy and electronic publication of thesis through the University of St Andrews.*

Date:

Signature of candidate:

Signature of supervisor:

CONTENTS

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Complexity is a young discipline which can help us understand the world around us. Self-organising emergent complex systems, in which large networks of components with no central control and simple rules of operation give rise to complex collective behaviour, are ubiquitous across disciplines [160]. In a cell, the complex network of chemical constituents and reactions play a key role in sustaining cellular functions [135]; our society is shaped by a network of repeated local interactions among individuals, which give rise to global regularities such as spontaneous formation of a common language and culture or the emergence of consensus about a specific issue [62]; even large-scale engineered systems, such as energy distribution networks [9], transportation networks [26] and the Internet [94], often evolve from the bottom up in a decentralised manner, growing complex structural patterns resulting from local optimisations and decision-making at different scales.

Although very different in nature, the systems mentioned above can all be abstracted into networks (or graphs), by describing which components in a system interact with one another, thus hiding the complexity of their constituting parts and concentrating on their connectivity patterns instead. The analysis of networks of interactions has proven useful in a wide range of systems and have recently become the focus of a dedicated research field called "network science". But the idea of abstracting complex problems away from their details into the language of graphs and then use general mathematical tools to manipulate them, is not new. It was originally invented in 1736 by one of the best-known mathematicians, Leonhard Euler, using which he solved the famous Königsberg Bridges problem (consisting in finding a round trip that traversed each of the bridges of the prussian city of Königsberg exactly once), and founded modern graph theory [44, 93]. Since its birth, graph theory has been established as a branch of discrete mathematics with stimulating problems such as graph colouring, covering and max-flow problems.

Recent years, however, have witnessed a substantial new movement in network research, with the focus shifting away from the analysis of networks small enough to be described in full, with all their nodes, edges and other details written down, to general statements about the properties of large-scale networks, whose complete details might not be known. It began with the introduction of random graphs by the Hungarian mathematicians Paul Erdős and Alfred Rényi, who studied the statistical properties of graphs generated by a random process [45, 89, 90]. Erdős and Rényi random graph model has guided our thinking about complex networks for decades since its introduction and is still widely used in many fields and serve as a benchmark for many modelling and empirical studies. But with the recent availability of large databases of various real networks, coupled with an increased computational ability to analyse them, the topology of real networks has been found to largely deviate from a random graph, calling for new tools and measurements to quantify their underlying organising principles, and leading the emergence of "network science".

Remarkably, it was discovered that the connectivity patterns of fundamentally different systems, including the World Wide Web [12] and the Internet [94], scientific coauthorship and citation networks [170], and neural and metabolic networks [55, 135], display common universal features, such as very large fluctuations in the number of connections – most of their nodes are very low connected, but there exists some nodes of very extreme connectivity (hubs). This is in contrast to random graphs where the majority of nodes have approximately the same number of connections. These discoveries have led to the development of new network formation models, based on various mechanisms such as preferential attachment and local optimisations to grow the irregular connectivity patterns found in many real networks, instead of random processes used to generate the random graphs that were used so far. Later, equipped with a better understanding of their complex structure, the focus has turned to investigate the dynamical behavior of networks, with a special emphasis on how the network structure affects the properties of a networked dynamical system. Indeed, nontrivial connectivity patterns were found to have tremendous dynamical implications, and network science has rapidly evolved and established itself as a leading scientific field in the description of complex systems [11, 42, 169].

But although the significant advances in our understanding of the structure and behaviour of complex systems, many challenges still remain in providing a comprehensive modelling framework that account for their important realistic features [122]. In particular, most studies so far have focused on isolated single networks that do not interact with or depend upon other networks, while in reality, networks rarely live in isolation and are often just one component in a much larger complex multilevel network. Examples include mutually dependent infrastructures, interactions between cells and cortical areas in our body, and people involved in more than one social network. Together with the increased availability of richer, bigger and multi-relational

datasets, in the last four years, coupled networks have been widely investigated and shown to give rise to a multitude of new features and phenomena that were not observed for isolated networks [103, 139].

In this thesis, we present analytical, numerical and empirical studies of coupled complex networks, aiming to understand the implications of coupling to the functionality and behaviour of complex systems, while tackling timely and important challenges in the network science literature. While in some cases there is an obvious separation into distinct coupled networks, in others, the mathematical advantage of multiple over single networks might not be evident. We demonstrate the opportunities that open up with the new degrees of freedom that result from the introduction of multiple layers, enabling the analytical study of heterogeneous multiscale networks, providing a clear and detailed picture into their behaviour that is often not possible to obtain within the mathematical framework of single networks. Throughout this thesis, we hope to maintain a stimulating discussion about what is expected to lead to a paradigm shift in the study of networks, moving from single to interacting networks, which is currently an extremely hot and active topic with new results published on a daily basis, attracting the attention of abundance of network scientists.

1.0.1 Contributions of this thesis

In the main part of this thesis, we present three theoretical models followed by an extensive empirical study of coupled transportation networks. The contributions of this thesis can be summarised as follows:

I A theoretical framework, based on generating functions, for studying the robustness of modular networks to attacks on interconnected and high betweenness centrality nodes.

II A theoretical framework, based on nonlinear ordinary differential equations, to study epidemic spreading in coupled adaptive networks, using which we discover a new equilibrium that was not seen before in single adaptive networks.

III A numerical study on the effect of nonrandom coupling on the behaviour of coupled networks, contradicting recent results on the topic, thus providing an important new insight.

IV An application of coupled networks approach to the study coupled transportation networks, providing the first empirical result on the interplay between the topologies of street and underground networks in metropolitan areas.

1.0.2 Structure

This thesis is organised as follows: in chapter 2 we review the main advances in the field of network science, providing a literature review while concentrating on the most relevant topics for the purpose of the present work. First, we review the qualitative measures used to characterise the large-scale organisation of complex networks, and the main network formation models, which aim to generate the structural features observed in real networks. Then we discuss in detail the two most widely used network dynamics models, percolation and epidemic spreading, while highlighting some of the striking mathematical results arising when "running" processes on networks with nontrivial connectivity patterns. We complete the chapter with a literature review on the recent advances made in the study of coupled networks.

In chapter 3 we present a theoretical framework for studying the robustness of modular or interconnected networks, exhibiting a multiscale structure consisting of tightly connected groups of nodes (models) with relatively few interconnections. We develop a simple model to generate modular networks with a varying number of modules and densities of interconnected nodes, and find a critical concentration of interconnections between modules, above which the internal structure of each module is inseparable from the system as a whole. We discuss the computer simulation used both to verify the analytical prediction, and to provide a more detailed picture of the dynamics, such as visualisation of the network at different stages and other statistical quantities not given by the analytical model. We conclude the chapter with a discussion about the implications of our results and possible directions in which they can be extended.

In chapter 4 we present a theoretical framework to study epidemic spreading in interconnected adaptive networks, which have the ability to adapt their topology dynamically in response to the dynamic states of nodes. We extend an existing analytical framework, based on nonlinear differential equations, from single to coupled adaptive networks, demonstrating the process of generalising "traditional" single networks results to account for the interaction and dependencies between networks. In particular, we discover a new equilibrium that only emerges in the case of weakly coupled networks, which in order to obtain its exact quantitative behavior from the analytical model, one must account for the actual second-order moments of the system, even for homogeneous networks, where such higher-order terms may generally be treated by a uniform approximation. Therefore, while tackling a specific problem, we demonstrate in this chapter that multilayer networks often require a more careful treatment than the one that is normally sufficient for isolated single networks.

In chapter 5 we present a numerical study on the effect of correlated coupling, accounting for correlations between the local properties of interconnected nodes, on the behaviour of coupled

networks, and in particular, how efficient correlated structures are in spreading flows. We present a model of constrained epidemic spreading, aiming to capture the resource constraints existing in coupled networks where, unlike connected nodes in a single network, coupled nodes often share resources, like time, energy, and memory. Using an extensive computer simulation, we analyse the model dynamics on networks with various topologies, revealing a qualitatively different result than the one obtained in recent studies, thus questioning their robustness, while also providing a possible explanation for the random coupling found in biological networks, which according to previous studies was considered less spreading-efficient. We complete the chapter with a discussion, calling for more future work about this topic, and especially more theoretical results.

To complement the theoretical work considered in the previous chapters, in chapter 6 we present a large-scale empirical study of interacting underground and street networks in the entire metropolitan areas of both London and New York. While intermodality was largely considered in the transportation science literature, most studies on the topic do not provide a topological analysis of the network's graph, a fact that has yet to be addressed in the complex networks literature. We aim to fill this gap in the literature while exploring the utility of coupled complex networks modelling, as well as demonstrating that they can deal with the scale and empirical complexity of real-world network exemplars.

Finally, we summarise our findings in chapter 7 comparing the different models presented in this thesis, and discussing current limitations and required future work both in the context on this thesis and in the more general context of coupled networks research.

CHAPTER TWO

BACKGROUND

Over the past fifteen years network scientists have developed many mathematical tools for understanding and predicting the behaviour of interconnected complex systems. These tools have found wide applicability to datasets taken from life sciences, social sciences, and physical sciences, as well as from engineered systems. In this chapter we review the main developments in this field, which are most relevant for the purpose of the present work. First, in section 2.1, we describe the mathematical structure of networks and review quantitative measures developed to characterise common properties observed in the topology of real networks. In section 2.2, we describe the main network formation models and their structural characteristics based on the measures presented in the previous section. In section 2.3 we discuss the most studied dynamical processes over networks, namely percolation and epidemic spreading, which will be repeatedly used throughout this thesis. Finally, in section 2.4, we review recent advances made in the study of coupled multilayer networks, which are the main focus of this thesis, before concluding in section 2.5.

2.1 Basic concepts in network theory

A network is an abstract representation of a set of entities or components in a system and the relations between them. The interconnected entities are called nodes or vertices and the relations between them are called edges or links. Mathematically, a network G, also called a graph, is defined by a pair of sets $G = (V, E)$ where V is a set of nodes and E is a set of links. The number of nodes in the network, also called the network size, is denoted by $N = |V|$. In a undirected graph, each of the links is defined by a couple of nodes $i, j \in V$, and is denoted by (i, j) meaning that the nodes i and j are connected, see Fig 2.1. In this thesis, as mostly the case in the complex network literature, we only deal with undirected simple networks which do not contain loops, i.e.

links from a node to itself, nor multiple links, i.e. couples of nodes connected by more than one link.

Both the nodes and links can be associated with additional variables, such as node states or link weights. For example, nodes could be assigned with one of two labels S or I representing susceptible and infected nodes in a network where epidemic is spreading. Links in this example could be assigned with a weight corresponding to the frequency of interactions which can pass a disease between the two nodes connected by the link. In the following we discuss some of the main statistical properties used to characterise the structure and behavior of networked systems.

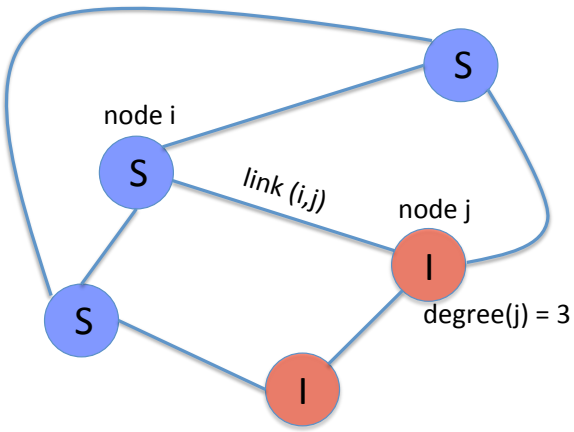


Figure 2.1: Schematic illustration of an undirected network with N=5 nodes, three of which are in S state (susceptible) and two are in I state (infected).

2.1.1 Degree distribution

The degree k_i of node $i \in V$ is its number of connections, see Fig 2.1. The most basic topological characterisation of a graph can be obtained in terms of the degree distribution, p_k, defined as the probability that a node chosen uniformly at random has degree k or, equivalently, as the fraction of nodes in the graph having degree k. The degree distribution offers a simple means to separate networks into classes. For example, many real networks, such as WWW [22, 49], collaboration networks [20, 170] and cellular networks [134, 135], have been shown to display power law degree distribution $p_k \sim k^{-\lambda}$ with a scaling exponent between $2 \leq \lambda \leq 3$. Thus, unlike homogeneous networks such as regular lattices or random graph, these networks, having a highly inhomogeneous degree distribution, result in the simultaneous presence of a few nodes (the hubs) linked to many other nodes, and a large number of poorly connected elements. Such networks have been named scale-free networks, because power-laws have the property of having the same functional form at all scales, see Fig. 2.2.

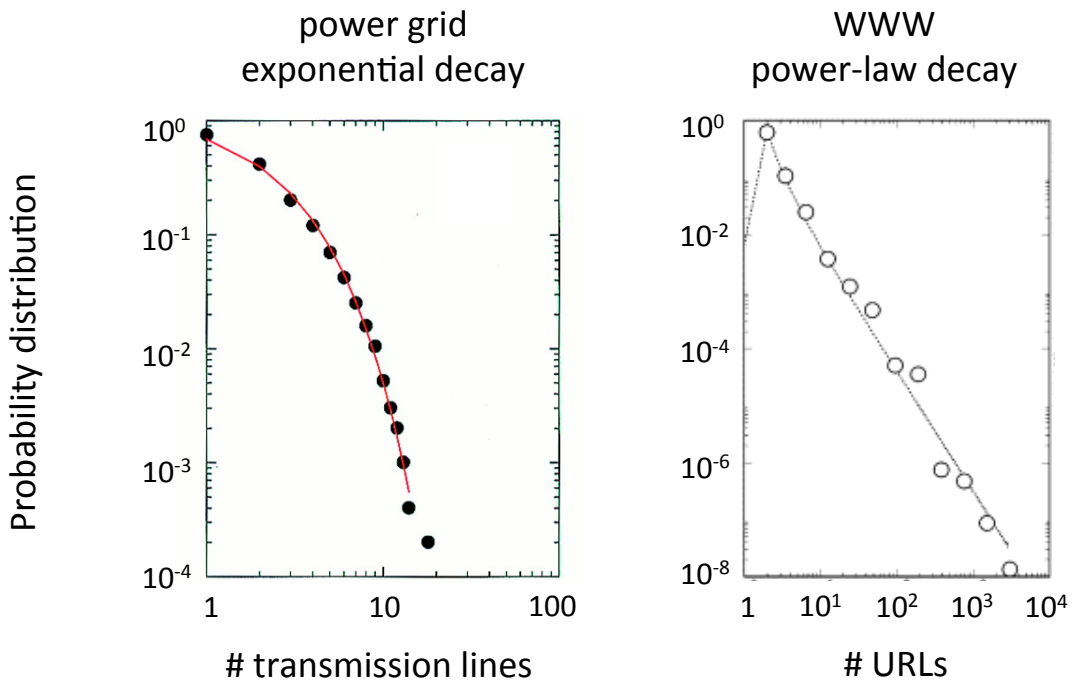


Figure 2.2: Degree distribution of the electric power grid of Southern California (figure adapted from [16]) and a network of documents connected by URLs obtained from the complete map of the nd.edu domain (figure adapted from [22]). Power-law distribution is characterised by a straight line in a log-log plot, thus also called scale-free distribution. The exponential decay observed in the power grid network is common in networks where constraints such as space and energy are preventing the formation of extremely high-degree nodes [16].

A very interesting property of scale-free networks, and even networks where the power law behaviour holds only in the tail, is that if the scaling exponent of the degree distribution is in the range $2 < \lambda \leq 3$, then in the limit of large network size ($N \to \infty$), its first moment (i.e. mean degree) is finite, while its second moment (related to the dispersion of the degree distribution) diverges [176]. Suppose we have degree distribution, p_k, that has a power-law tail for $k \geq k_{\min}$. Then, its mth moment is given by

$$\begin{aligned}\langle k^m \rangle &= \sum_{k=0}^{k_{\min}-1} k^m p_k + \underbrace{C \sum_{k=k_{\min}}^{\infty} k^{m-\lambda}}_{p_k = Ck^{-\lambda}} \\ &\simeq \sum_{k=0}^{k_{\min}-1} k^m p_k + C \int_{k_{\min}}^{\infty} k^{m-\lambda} \mathrm{d}k\end{aligned}$$

$$= \sum_{k=0}^{k_{\min}-1} k^m p_k + \frac{C}{m-\lambda+1}[k^{m-\lambda+1}]_{k_{\min}}^{\infty} \tag{2.1}$$

The first term is a finite number and the second term diverges for $m-\lambda+1 \geq 0$. Thus, for $2 < \lambda \leq 3$, we obtain that the first moment is finite and the second moment diverges.

2.1.2 The small-world effect

The *distance* between two nodes $i, j \in V$, d_{ij}, is defined as the number of links in the shortest path from one node to the other (if there is any). The maximum value of d_{ij} is called the *diameter* of the graph, denoted by $Diam(G)$. A measure of the typical separation between two nodes in the graph is given by the average shortest path length, also known as *characteristic path length*, defined as the mean distance over all couples of nodes

$$L = \frac{1}{N(N-1)} \sum_{i,j \in V, i \neq j} d_{ij}. \tag{2.2}$$

In regular d-dimensional lattices, the average path length grows with the lattice size as $L \sim N^{\frac{1}{d}}$ [42]. In contrast, in most of the real networks, despite their often large size, there is a relatively short path between any two nodes. This property is known as the "small-world" effect and is mathematically characterized by slow scaling $L \sim \ln N$ [241]. Another related notation is the "six degrees of separation", which refers to Milgram's experiments, where a path of first-name acquaintances with a typical length of about six was found between most pairs of people in the United States [157, 240].

But, although found in various types of real networks [241], including biological and technological ones, the small-world effect does not imply a particular organisation principle, rather it is an obvious mathematical property in some network models, including totally random networks obtained by randomly placing links among a given number of nodes, which will be introduced in section 2.2.1. Scale-free networks are considered "ultra-small", since their characteristic path length scales even slower with the system size $L \sim \ln\ln N$ [68].

2.1.3 Clustering

In contrast to random graphs, the small-world property in real networks is often associated with the presence of *clustering*, meaning that in social networks, for instance, two friends of someone are often also friends with each other. This is property can be quantified by the *clustering coefficient* defined as the number of links among the neighbours of node i, normalised by the

possible number of links [241]

$$C_i = \frac{2E_i}{k_i(k_i-1)} \tag{2.3}$$

where k_i is the degree of node i (i.e. number of friends), $\frac{k_i(k_i-1)}{2}$ is the maximum number of links possible among the neighbours of node i, and E_i counts how many of these links actually exists (i.e. how many friends of node i are also friends with each other). The clustering coefficient of the whole network is the average of all individual C_i's.

An alternative measure, called *transitivity* measures the fraction of triples that have their third link filled in to complete the triangle

$$T = \frac{\text{3 x number of triangles in the network}}{\text{number of connected triples of nodes}} \tag{2.4}$$

In other words, T measures ratio of the means, rather than the mean of ratios, giving less weight to contributions of low-degree nodes [169].

Regardless of which definition, most real networks exhibit a significant amount of clustering compared to random networks with the same number of nodes and links [169]. Moreover, it is suspected than real networks have a nonzero clustering limit when the network becomes infinitely large, so that $C = O(1)$ as $N \to \infty$, in contrast to random networks where $C = N^{-1}$ [10].

2.1.4 Giant component

A connected component of a network is a maximal subset of nodes that are connected by paths through the network. The size of the largest connected component is an important quantity, and in the limit of large network size ($N \to \infty$) is equated with the *giant component* which contains a constant fraction of nodes of an infinite network. The existence of a giant component indicates that most nodes in the networks, $O(N)$, are reachable from one another, such that a rumour, for example, could spread from one person to almost everyone else, unlike in networks composed of many small components, not connected to one another (i.e. no giant component exists), where the rumour could only spread within each component separately without invading the network "globally". Since one isolated node with no links is already enough to make a network not connected (a connected network is one in which all nodes are connected to one another), the existence of a giant component is an important statistical property used to measure connectivity in a network.

2.2 Network models

In this section we present some of the main network formation models aim at generating specific topologies that reproduce observed statistical features of real-world networks.

2.2.1 The Erdős-Rényi model

The classical Erdős-Rényi (ER) model of random graphs generates a random undirected network with N labeled nodes connected by M edges, which are chosen randomly from the $N(N-1)/2$ possible edges [89, 90]. The random graph ensemble with exactly N nodes and M links is sometimes denoted as $\mathscr{G}(N,M)$, forming a probability space in which every realisation out of the possible $C^M_{N(N-1)/2}$ is equiprobable.

An alternative and equivalent definition of a random graph, which is easier to analyse mathematically, is the binomial model, $\mathscr{G}(N,p)$, where every pair of nodes being connected with probability p. Since the presence or absence of edges is independent, the resulting network has Poisson degree distribution with mean $z = p(N-1)$

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \simeq \frac{z^k e^{-z}}{k!} \tag{2.5}$$

with the last approximate equality becoming exact in the thermodynamic limit of large N and fixed z [169].

The structural properties of ER random graphs vary as a function of p, showing, in particular, a phase transition at a critical probability $p_c = \frac{1}{N}$ above which an extensive (i.e., $O(N)$) fraction of all nodes are joined together in a single giant component (see section 2.1.4), and the rest of the nodes occupying smaller components with exponential size distribution and finite mean size [42, 169]. This result will be discussed in detail in section 2.3.1, where we present site and bond percolation processes.

ER random graphs are small-worlds - almost all graphs with the same N and p have precisely the same diameter concentrated around $Diam = \frac{\ln N}{\ln z}$, and characteristic path length behaves the same $L \sim \frac{\ln N}{\ln z}$ [11]. However, in almost all other respects, random graphs do not reproduce typical features of real-world networks such as a scale-free degree distribution or strong clustering. Due to the independent placement of links, they have clustering coefficient $C = p$ which tends to zero as N^{-1} in the limit of large system size. Nonetheless, ER random graphs are an attractive baseline model for various applications. Their simplicity makes them easy to analyse mathematically, and more importantly, they constitute the basic building blocks of network theory, and our understanding about the way networks behave.

2.2.2 Watts-Strogatz small-world networks

In order to capture the observed strong clustering in real networks together with short path lengths, Watts and Strogatz (WS) generated a network interpolating between a regular graph and a random graph [241], see Fig 2.3. Starting with a ring lattice with N nodes in which every node is symmetrically connected to its m ($m/2$ on either side) nearest neighbors for a total of $M = \frac{mN}{2}$ edges, create shortcuts by randomly rewire each link of the lattice with probability p such that self-connections and duplicate edges are excluded.

For $p = 0$ we have a regular lattice with characteristic path length $L(0) \simeq \frac{N}{2m} \gg 1$ and clustering coefficient $C(0) \simeq \frac{3}{4}$. On the other hand, for $p = 1$ the model produces a random graph with $L(1) \sim \frac{\ln N}{\ln m}$ and $C(1) \sim \frac{m}{N}$. For intermediate values of p between these limits, Watts and Strogatz found an interval where $L(p)$ is close to $L(1)$ yet $C(p) \gg C(1)$, thus producing a network with both the small-world property and a non-trivial clustering coefficient [11].

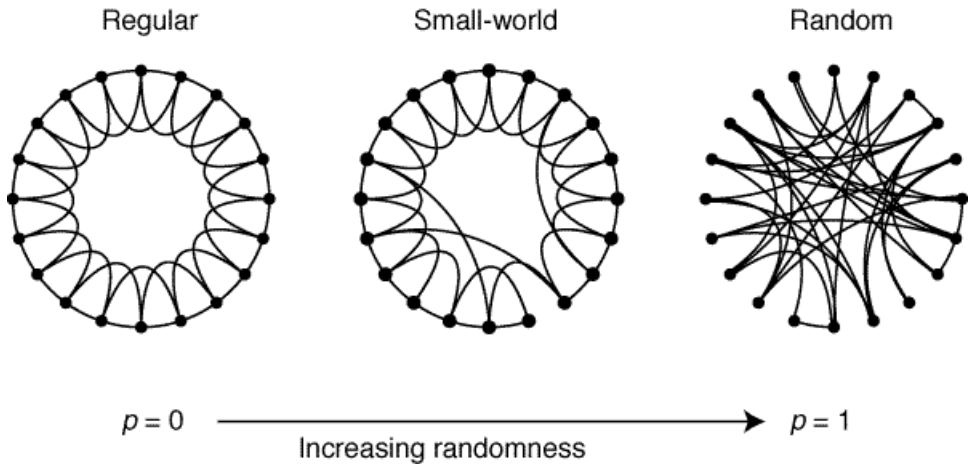


Figure 2.3: Construction of the Watts-Strogatz small-world model: interpolation between a regular graph ($p = 0$) and a random graph ($p = 1$). Watts and Strogatz showed that for intermediate values of p, the obtained network has both a small characteristic path length and a high clustering coefficient. Figure adapted from [241].

2.2.3 Barabási-Albert preferential attachment model

Both the ER and the WS model discussed above produce networks with narrow degree distributions (see Fig. 2.4), unlike most real-world networks (see section 2.1.1). The Barabási-Albert (BA) model [21] overcomes this flaw while attempting to provide an explanation to the origin of the highly skewed degree distributions of real-world networks. The model incorporates two essential ingredients, growth and preferential attachment, to produce scale-free networks. Starting with a small number of nodes, m_0, at every time step, a new node with $m \leq m_0$ links

is added to the network. The new node connects to nodes already present in the network with probability proportional to their degree, i.e. according to linear preferential attachment. This model has been shown to produce a network with degree distribution $p_k \sim k^{-3}$ in the limit of large networks [11]. The characteristic path length is smaller than in ER networks with the same number of nodes and links [11], and has been shown to scale with the number of nodes $L \sim \frac{\ln N}{\ln \ln N}$[45]. The clustering coefficient decays slower than in random graphs $C \sim N^{-0.75}$, but still vanishes with the system size.

Substituting the linear preferential attachment with sub-linear, super-linear or constant attachment probability, has been shown to produce networks which are no longer scale-free [141], thus suggesting that linear preferential attachment is essential to the formation of scale-free networks. But Barabási and Albert were not the first to point this out. Already in 1965, Derek de Solla Price found that citation networks have power-law degree distributions [79], and consequently developed a model where the probability that a new published paper cites a previous one is taken to be proportional to $k_{\text{in}} + 1$, where k_{in} is the number of times that the paper has already been cited [190]. Price's model itself is built on ideas developed in the 1950s by Herbert Simon, who showed that power laws, which appear in a wide range of empirical data, arise when "the rich get richer" [212]. However, Barabási and Albert were the first to realise the enormous potential of the model and its relevance for a wide range of real-world networks. Moreover, having neglected the direction of links and fixing the number of links added with each new node, the BA model is simpler than the one proposed by Price and is thus more intuitive and attractive. For that reason, the preferential attachment model is usually referred to as the BA model.

2.2.4 Configuration model

The configuration model, introduced by Bender and Canfield [36], is a generalisation of the ER random graph model, which allows to generate graphs with arbitrary degree distributions. Given a degree sequence $D = \{k_1, k_2, \ldots k_N\}$, each node i is attached with k_i "stubs" sticking out of it, which are the ends of links-to-be. Then, the graph is constructed by randomly choosing pairs of stubs and connecting them together. This procedure generates, with equal probability, every possible topology of a graph with the given degree sequence [161]. D is chosen in such a way that the fraction of nodes with degree k will tend to the desired degree distribution for large N, for example, by drawing a random sample from the degree distribution. The obtained graphs may have loops and multiple links, but these can be neglected or discarded in the large network limit, $N \to \infty$.

The configuration model is widely used to generate random uncorrelated networks, especially with scale-free degree distributions, see Fig. 2.4. Its simplicity makes it an attractive model to

study analytically. For example, the exact condition for the existence of a giant component and its expected size was provided by Molloy and Reed [161, 162]. Newman et al. [175] studied the average size of non-giant components, and Chung and Lu [132] studied the average distances between nodes.

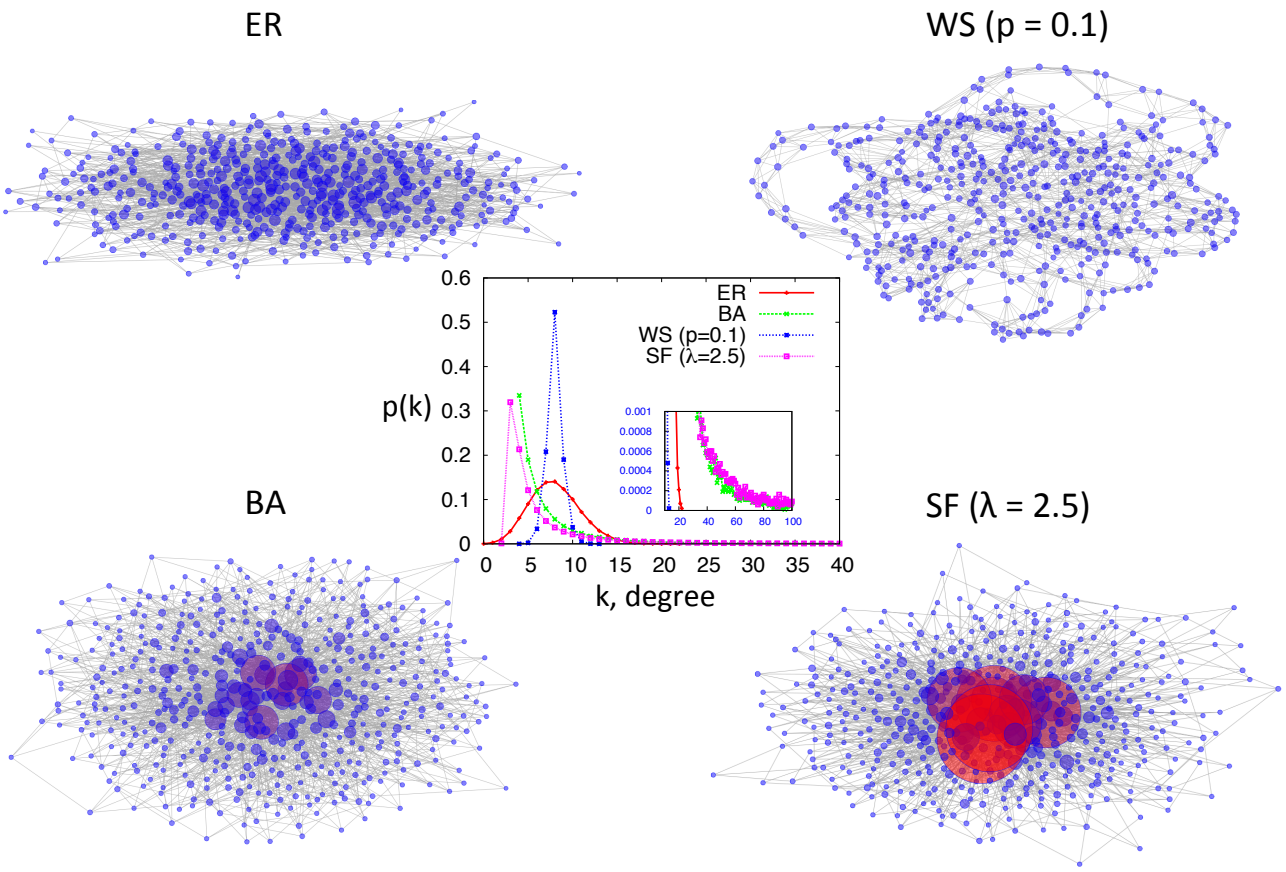


Figure 2.4: Visualisation of random networks with $N = 500$ nodes and mean degree $\langle k \rangle = 8$, obtained through the ER model, WS model with $p = 0.1$, BA model, and the configuration model using a degree sequence drawn from a scale-free degree distribution with scaling exponent $\lambda = 2.5$. Nodes sizes and colours correspond their degree. The graph in the middle shows the obtained degree distributions for larger graphs ($N = 100\,000$) generated in the same way. As expected, the degree of nodes in the WS network, which is based on a regular-degree lattice, is most narrowly distributed (all nodes have almost exactly the same colour and size), with a maximum node degree in the visualised network equal to 11 (only 3 more than the average 8). The ER network displays slightly more fluctuations in degree (we can see some larger nodes in the core of the network), but the degree distribution has a very short tail (see inset). However, the BA and scale-free networks both have heavy-tailed degree distributions (see inset), with extremely well connected nodes, especially in the scale-free network since its scaling exponent is 2.5, where for the BA it is 3.

2.3 Dynamics on networks

As described in the previous two sections, network science initially revolved around developing quantitative tools to characterise the structure of real-systems, following by the development of network formation models which generate their observed statistical features. With the advances in our understanding of their complex structure, the focus of network science has turned to investigate the implications of such structures to the behaviour and functionality of networked dynamical system. This still remains the ultimate goal and thus the main focus of network science, and in the following, we describe the two most important and widely used models of networks dynamics, which are also repeatedly used throughout this thesis.

2.3.1 Percolation and network resilience

One of the first examples to be studied thoroughly of a process taking place on a network has been site and bond percolation processes. A site (bond) percolation process is one in which nodes (links) are removed from the graph with *failure probability* f. The probability to remain in the graph, $p = 1 - f$, is often called the *occupation probability*, since we could think of removed nodes (links) as "unoccupied" (i.e. nonfunctional or nonoperational), and the remaining nodes (links) as "occupied".

When the occupation probability is small, the network is composed of a large number of very small connected components, unreachable from each other through occupied nodes. But as the occupation probability increases, it reaches a critical value, p_c, called the *percolation threshold*, above which a giant component emerges, connecting a positive finite fraction of the nodes in the network [46]. The percolation threshold can be used as a measure for network resilience - the smaller it is the more robust is the network since a small fraction of occupied nodes is already enough for the giant component to emerge, i.e. to have global connectivity.

One of the most interesting results in the study of complex networks is that infinitely large networks with power-law degree distributions $p_k \sim k^{-\alpha}$ for some constant $2 \leq \alpha \leq 3$ have a critical value $p_c = 0$, indicating that the network always has a giant component, or in the language of physics, the network always percolates [67]. To gain a better understanding of this striking results, in the following we present the generating functions approach developed by Newman *et al.* [175] to analytically study a site percolation process on random networks.

2.3.1.1 Generating functions

Consider a random graph with a large number of nodes, N, and structure of the configuration model described in section 2.2.4, i.e. nodes are randomly assigned with degrees independent

identically distributed from some degree distribution p_k, and are randomly connected with the only constraint that a node with degree k has exactly k links. The generating function of the degree distribution p_k is a power series whose coefficients are the degree probabilities $p_1, p_2, \ldots$

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \tag{2.6}$$

where $|x| \leq 1$, and the distribution p_k is assumed correctly normalized, so that $G_0(1) = 1$.

Generating functions are an extremely powerful tool in combinatorial enumeration problems allowing us to use functional manipulations to study, for example, the average number of nodes in a network component [244]. In the following we review some of their properties that will prove useful in subsequent development, and especially in chapter 3. First, we observe that the generating function G_0 given in 2.6 indeed "generates" the probability distribution p_k, meaning that given G_0, we can retrieve the component probabilities of p_k. The probability that a node has k connections is given by the kth derivative of G_0 according to

$$p_k = \frac{1}{k!} \frac{\mathrm{d}^k G_0}{\mathrm{d}x^k}|_{x=0} \tag{2.7}$$

The next useful property of the generating function is that we can use it to extract summary statistics directly. The mean of the distribution, which corresponds to the mean degree, $\langle k \rangle$ is given by

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k = G_0'(1) \tag{2.8}$$

Higher moments of the distribution can be calculated from higher derivatives also, and in general, we have

$$\langle k^n \rangle = \sum_{k=0}^{\infty} k^n p_k = \left[\left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^n G_0(x) \right]_{x=0} \tag{2.9}$$

Finally, generating functions have a "powers" property, that if the distribution of a property α of an object is generated by a given generating function, then the distribution of the total of α summed over m independent realizations of the object is generated by the mth power of that generating function. For example, the distribution of the sum of the degrees of m nodes chosen at random from a network with degree distribution p_k, is generated by $\left[G_0(k)\right]^m$ [175].

We continue by defining the generating function for the excess degree distribution defined as follows. Consider the probability of following a randomly selected link to reach a node with an additional k links, apart from the one by which it has been reached. This probability is proportional to the degree, $k+1$, of the node since there are $k+1$ links that arriving at this node

(which makes it $k+1$ times more likely to be arrived at than a 1 degree node). Therefore, the distribution of the remaining degrees of nodes reached by following a random link, called the *excess degree*, is $q(k) = \frac{(k+1)p(k+1)}{\sum_{k=1}^{\infty} kp(k)}$ and the associated generating function is given by

$$G_1(x) = \frac{\sum_{k=0}^{\infty}(k+1)p_{k+1}x^k}{\langle k\rangle} = \frac{1}{\langle k\rangle}G_0{}'(x). \tag{2.10}$$

2.3.1.2 Connected components

Let us now consider the distribution of the sizes of connected components in the graph. Let $H_1(x)$ be the generating function for the distribution of the sizes of components that are reached by choosing a random link and following it to one of its ends. Note that the giant component, if there is one, is excluded from $H_1(x)$. Thus, component sizes are finite and the chances of a component containing a closed loop of links goes as N^{-1} which can be neglected in the limit of large N. In other words, each component is treelike in structure, consisting of the single node v, we reach by following our initial link, plus any number plus (including zero) of other treelike components, with the same size distribution, joined to it by single links. Summing over all the types of connectivity possible for v leads to the self consistency equation for $H_1(x)$

$$H_1(x) = \underbrace{x\,q_0}_{\text{node } v \text{ has no additional links}} + \underbrace{x\,q_1\,H_1(x)}_{\text{node } v \text{ has 1 additional link}} + \underbrace{x\,q_2\,[H_1(x)]^2}_{\text{node } v \text{ has 2 additional link}} + \ldots \tag{2.11}$$

However, q_k is nothing other than the coefficient of x^k in the generating function $G_1(x)$, see equation 2.10, and hence equation 2.11 can also be written

$$H_1(x) = xG_1(H_1(x)). \tag{2.12}$$

Starting from a randomly chosen node, rather than link, we have one such component at the end of each link leaving that node, and hence the generating function for the size of the whole component is

$$H_0(x) = xG_0(H_1(x)). \tag{2.13}$$

Using the equations developed above we can now find properties of interest such as the mean component size to which a randomly chosen node belongs. According to Eq. 2.8, this quantity can be extracted from the generating function for the distribution of components sizes according to

$$\langle s\rangle = H_0{}'(1) = 1 + G_0{}'(1)H_1{}'(1). \tag{2.14}$$

From equation 2.12 we have

$$H_1{}'(1) = 1 + G_1{}'(1) H_1{}'(1) \tag{2.15}$$

and by substituting into equation 2.14 we obtain

$$\langle s \rangle = 1 + \frac{G_0{}'(1)}{1 - G_1{}'(1)}. \tag{2.16}$$

The point where the average component diverges, $G_1{}'(1) = 1$, marks the phase transition at which a giant component first emerges. This yields a critical occupation probability, above which the giant component exists, $p_c = \frac{1}{G_1{}'(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle}$, where $\langle k^2 \rangle$ is the second moment of the degree distribution p_k. Therefore, in scale-free networks where the first moment is finite and second moment diverges in the limit of large network (see equation 2.1), p_c goes to zero with the network size, i.e., the network contains a giant component when any finite fraction of the nodes or links are removed.

At the point where the giant component first emerges, $H_0(1)$ is no longer unity (since it excludes the giant component) but is equal to $1 - S$, where S is the fraction of nodes in the giant component. Thus, we obtain

$$S = 1 - H_0(u) = 1 - G_0(H_1(u)) \tag{2.17}$$

where $u \equiv H_1(1)$ is the smallest non-negative real solution of

$$G_1(u) = u \tag{2.18}$$

since by definition $u = H_1(1) = G_1(H_1(1)) = G_1(u)$.

2.3.1.3 Attack tolerance

Finally, we would like to briefly discuss the extension of Callaway *et al.* [57] to the approach presented here. In their method, the occupation probability is no longer fixed across the network, but instead it is a function of the degree of nodes. Let r_k be the occupation probability of a node with degree k. We define the generating functions for the distribution of the degree and excess degree of occupied nodes:

$$F_0(x) = \sum_{k=0}^{\infty} p_k \, r_k \, x^k, \quad F_1(x) = \frac{F_0'(x)}{\langle k \rangle} \tag{2.19}$$

Then, the probability distribution of the size of the component of occupied nodes to which a randomly chosen node belongs is generated by $J_0(x)$, where

$$J_0(x) = 1 - F_0(1) + xF_0(J_1(x)), \quad J_1(x) = 1 - F_1(1) + xF_1(J_1(x)) \tag{2.20}$$

and the mean component size is given by

$$\langle s \rangle = F_0(1) + \frac{F_0'(1)F_1(1)}{1 - F_1'(1)}. \tag{2.21}$$

Once the giant component emerges, which happens at $F_1'(1) = 1$, its size is given by

$$S = F_0(1) - F_0(u), \quad u = 1 - F_1(1) + F_1(u) \tag{2.22}$$

Using this approach, Callaway *et al.* [57] studied the resilience of networks to attack in which nodes are removed in order from highest degree to lowest degree, and confirmed previous numerical studies showing that scale-free networks are highly susceptible to this kind of attack, much more than networks with narrow degree distributions such as ER networks [13, 49], see Fig. 2.5. We further demonstrate an application of this approach to study attacks in multiple networks in chapter 3.

2.3.2 Epidemic spreading

The second set of processes that have been widely investigated, together with bond and site percolation described above, is epidemiological processes. Here we are interested in what are the network topologies that give rise to disease, information, and rumour spreading over them. We begin by describing the two most studied models in epidemiology, susceptible-infected-recovered (SIR) and susceptible-infected-susceptible (SIS) models of epidemic disease [17, 19, 124]. The SIR model describes diseases resulting from permanent immunisation or death of infected individuals. In this model, each individual can be in one of three possible states: (S) susceptible corresponds to healthy individuals who do not have the disease but can catch it if exposed to infected individuals, (I) infected corresponds to individuals who have the disease and can pass it on, and (R) recovered or removed corresponds to individuals who recovered from disease and have permanent immunity, or individuals who died from the disease, and they can never get it again or pass it on. The SIR model has been shown to describe diseases such as chickenpox, HPV, seasonal influenza, and H1N1 [17, 70], as well as rumour-spreading [164, 247] and the spread of computer viruses [150]. It is based on two parameters, infection rate, β (i.e. infection probability per unit time) in which susceptibles can catch the disease from infected, and recovery

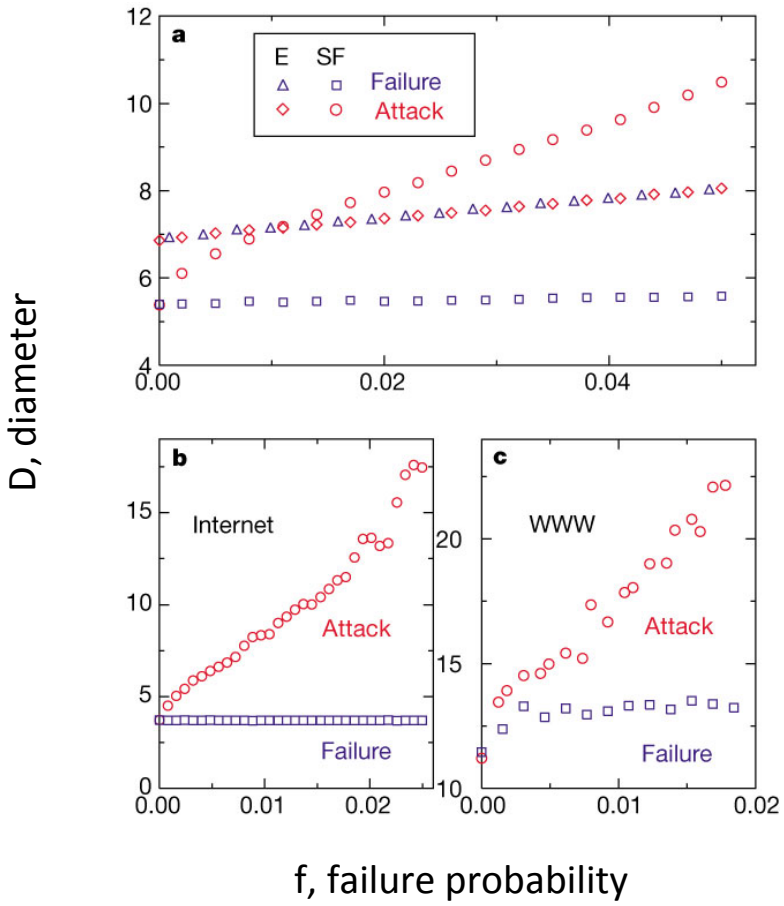


Figure 2.5: Robustness of scale-free (SF) and exponential (E) networks to attack and random failure. As demonstrated using both random networks models (a) and real-world networks (b)-(c), scale-free networks are extremely resilient to random nodes failure, where the network diameter is almost unaffected in the range shown. However, the diameter of scale-free networks is increasing very rapidly when attacking nodes in order from highest degree to lowest degree, much more than exponential networks. Figure adapted from [13].

rate, γ, in which infected recover and become immune to the disease. Under the *fully mixed assumption*, where one assumes that any susceptible can catch the disease from any infected (i.e. individuals they make contact with are randomly chosen from the whole populations), the time evolution of the disease is given by

$$\frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i \tag{2.23}$$

where s, i, and r is the fraction of susceptible, infected and recovered individuals respectively. In other words, s, i, and r is the probability that a node is susceptible, infected or recovered respectively, and therefore multiplying generates the compound probability of having both in any pair chosen at random, with full mixing guaranteeing no bias in terms of node adjacencies.

In some spreading processes, individuals (or other entities such as computers) can catch the same disease more than once, and thus in the SIS model, infected individuals recover in rate γ and become susceptible again, describing diseases such as tuberculosis, gonorrhea and common cold [17, 19, 124], as well as computer viruses in systems with no automatic updated antivirus programs [183, 184]. Again, under the fully mixed assumption, the time evolution of a SIS spreading process is given by

$$\frac{ds}{dt} = -\beta is + \gamma i, \quad \frac{di}{dt} = \beta is - \gamma i \tag{2.24}$$

The most important point is that the dynamical equations of both SIR and SIS always yields a non-zero epidemic threshold (also called basic reproduction number) corresponds to critical infection rate above which the disease persists (in case of SIS), or the disease spreads and infects a non-zero fraction of the population in the limit of large system size (in case of SIR).

But unlike the fully mixed assumption, in reality, diseases can only spread between people who have physical contact. Therefore it is important to consider the structure of the contact network, especially in cases like sexually transmitted diseases where there is a large heterogeneity in degrees of sexual activity within the overall population [17, 19, 124]. When a disease is spreading over a network, individuals, represented by nodes, can only catch the disease from their network neighbours. Thus, moving from differential equations to processes over a network allows us to consider the structure of interactions, which is otherwise neglected by the fully mixed assumption. Grassberger showed that SIR spreading over a network can be mapped exactly onto bond percolation on the same network [111]. Then, the percolation threshold corresponds to the epidemic threshold; the distribution of percolation clusters (i.e., components connected by occupied links) corresponds to the distribution of the sizes of disease outbreaks that start with a randomly infected node; and the size of the giant component corresponds to the size of the epidemic outbreak. Therefore, the result discussed in section 2.3.1, where scale-free networks with scaling exponent $2 \leq \lambda \leq 3$ always percolate, means that in these networks there is no non-zero epidemic threshold. In other words, diseases will always propagate in these networks, regardless of the infection and recovery rates [153].

Although the SIS model cannot be solved exactly on a network as can the SIR model, Pastor-Satorras and Vespignani provided a mean-field approximation based on differential

equations [183, 184]. In short, the idea is to allow the rate of infection to vary among nodes based on their degree, replacing i and s with i_k and s_k representing the fraction of nodes of degree k that are infected or susceptible. The advantage of this approach is that it can tell not only the long time behaviour of the outbreak (e.g. finial outbreak size, critical infection rate), but also the time evolution of an outbreak. Pastor-Satorras and Vespignani showed that also in the SIS case, where the network has power-law degree distribution with scaling exponent $2 < \lambda \leq 3$, the disease will never die out regardless of the infection and recovery rates.

Obviously, real systems are always finite and thus even in scale-free networks there is always an effective non-zero epidemic threshold, below which the epidemic will not spread. In addition, scale-free networks with high clustering coefficients [88], as well as networks embedded in regular Euclidean space [197, 237], have also been shown to exhibit non-zero epidemic thresholds. But although the absence of non-zero epidemic threshold does not hold in various types of systems, scale-free networks are still proved to be very good spreaders. For example, it has been shown that the epidemic threshold scales with system size as $\beta(N) \sim \frac{1}{\ln N}$ and is therefore very small even for not very large networks, and is significantly smaller than that of a random graph with the same size [185]. Therefore, various immunisation strategies on scale-free networks have been examined [17, 81, 186]. Uniform immunisation has been shown to be totally ineffective on scale-free networks, since it depresses the infection prevalence too slowly, and thus, in the limit of infinite system size, the critical fraction of nodes to immune such that an epidemic will not break is one, i.e. all nodes need to be vaccinated. But even for finite systems, unrealistically high densities of randomly immunised individuals are required to stop the epidemic from breaking. However, targeted immunisation based on the node connectivity is highly effective and can potentially eradicate a virus.

And fortunately, there is even an easy way to find high-degree nodes without global knowledge of the network, which is very important in cases like sexual contact networks, where it is hard to obtain data. Cohen *et al.* [69] have pointed out that the probability of reaching a particular node by following a randomly chosen link is proportional to the nodes degree (see excess degree in section 2.3.1). Therefore, by choosing a random person from a population and vaccinating one of their friends, an efficient targeted immunisation can be achieved. And in fact, the "contact tracing" method, that has been widely effective in controlling STDs [65, 95] and SARS [149, 193], is taking a very similar approach.

2.4 Coupled networks

The exploding body of work that has been described in previous sections is providing a firm basis to the evolution of network science. The major challenge at this point is to account for more realistic features of networks such as strong coupling between networks (networks are not isolated), the time evolution of networks (networks are not static), the coevolution of structure and dynamics (structure is affected by dynamics), other classes of links including different signs of interactions, and spatial properties including geographical aspects of networks [122].

And indeed, in recent years there has been a great effort to tackle some of these challenges. For example, while most of the studied until now assumed that networks are not changing over time, or that the time-scale of structural changes is much slower than the dynamics taking place, the emerging topic of temporal networks is examining networks where links are generated, disappear and reappear over time, developing methods for analysing topological and temporal structure and models for explaining their relation to the behavior of dynamical systems [126].

Adaptive networks, combining dynamics on a network with dynamical adaptive changes of the underlying network topology have only recently started to get attention [115, 116]. Examples include social networks where links with ill people (dynamics) might be temporarily removed (structure) in order to avoid infection, or repeating traffic congestions (dynamics) on a given road could lead to formation of new roads (structure). In these networks, a feedback loop between the state and topology of the network is formed, giving rise to a remarkably rich dynamical behaviors, which are very difficult to study analytically. We will expand on this topic in chapter 4.

Different types of interactions between entities in a system, such as negative and positive social links [147], as well as dependency links representing strong local relations [181], have been recently suggested explaining social phenomena and dynamics that could not be addressed before.

Finally, research on spatial networks aims to understand how the spatial constraints affect the structure and properties of networks which are embedded in space, such as the Internet, mobile phone networks, power grids, and neural networks [26].

But undoubtedly, the topic that has been getting most attention since the revolutionary paper of Buldyrev *et al.* [53] that was published in 2010, is coupled multilevel networks. Both natural and engineered systems are rarely isolated. They interact with the environment and one another constituting a part of a much larger complex multilevel system. Such examples include people involved in more than one social network [238], proteins in a cell interacting with other proteins through both Protein-Protein Interaction (PPI) and metabolic pathways [23],

and critical infrastructures depending both on one another [195] and on political and social processes [61]. As technology has advanced, the coupling between individual networks, that were treated as isolated systems until now, is becoming stronger and stronger. Aside from physical, logical and geographical interdependencies, most systems today, including critical infrastructures such as power-grids, exhibit cyber-dependency, where normal functioning relies on information technology [194, 195, 234].

And the reason that these strong couplings and interdependencies are attracting so much attention, is that they seem to give rise to completely new phenomena, that have never been observed in isolated single networks. Havlin *et al.* [122] compares this paradigm shift from single networks to coupled multilevel networks, to the interactions between particles in physics: "As in physics, when only the individual particles were studied it was made possible to understand the properties of gas; however, when the transition was made to study the interactions between these particles, it was finally made possible to understand and describe liquids and solids. Thus, such a transition in network science will lead to a significant paradigm shift, which will reveal a multitude of new features and phenomena."

2.4.1 Interdependent networks

In an attempt to provide a mathematical framework for analysing the consequences of cascades of failures occurring in interdependent critical infrastructures [194, 195], Buldyrev *et al.* [53] defined a new class of networks called *interdependent networks*, and studied a percolation process in a system of two interdependent networks. In these networks, unlike the connectivity links within each networks, the networks are interconnected by *dependency links*, representing the fact that the function of a given node in one network depends crucially on nodes in other networks. Thus, when a node from one network is removed in a percolation process, its dependent node from the other network is automatically removed as well.

This model was designed to capture the situation observed in real-world data from a power network and an Internet network (a supervisory control and data acquisition system). The data was extracted from an electrical blackout that affected much of Italy in 2003, where the shutdown of power stations directly led to the failure of nodes in the Internet communication network (since switches rely on electricity), which in turn caused further breakdown of power stations (since they rely on the Internet for control and recovery) [195].

In their model, Buldyrev *et al.* consider two equal size networks with arbitrary degree distribution, where each node from one network is mutually dependent on a randomly selected node from the other network. Using a generating function approach, they show that the resulting networks

are significantly more vulnerable than their non-interacting counterparts, where small failure in one network may lead to catastrophic consequences that breaks the whole system. This behavior is characteristic of a first-order phase transition, in contrast to the second-order phase transition characterising percolation of a single network, where the size of the giant component is decreasing *continuously* as the number of removed nodes increases, see Fig. 2.6. Here, instead, when a critical number of nodes are removed, the giant component suddenly (i.e. *discontinuously*) collapses, resulting in a first order percolation transition. Perhaps more importantly that the qualitative change in the percolation transition, the authors show that in contrast to single networks, interdependent networks with broader degree distributions are more vulnerable. Specifically, the result where single scale-free networks always percolate in the limit of large networks (see section 2.3.1), is no longer valid for interdependent scale-free networks.

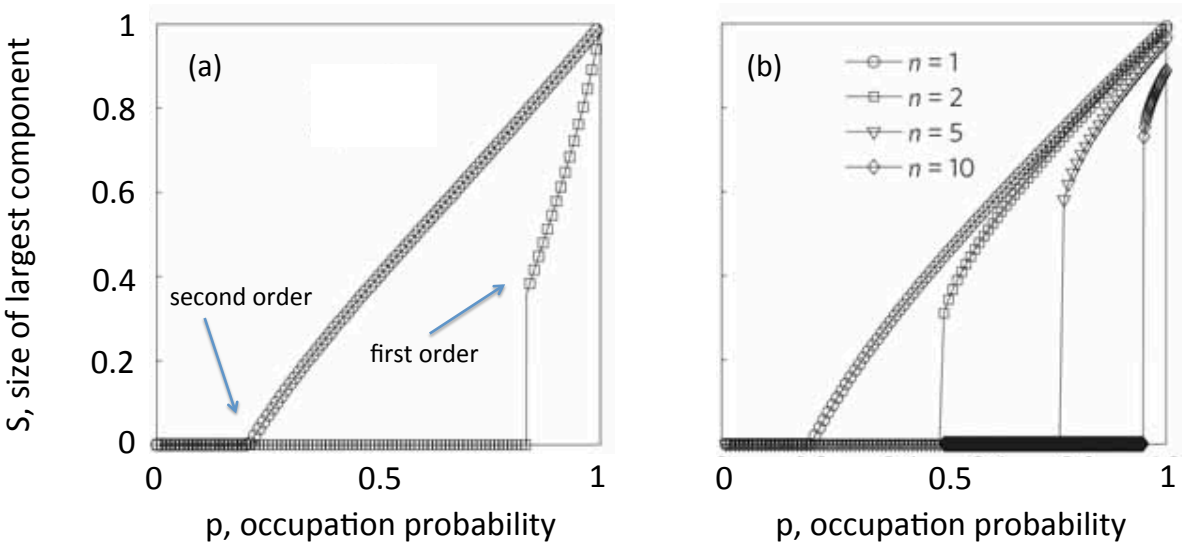


Figure 2.6: Second order vs. first order percolation transition. (a) In two weakly interdependent networks (with few interdependency links between them), the size (fraction of nodes) of the largest connected component is changing continuously with the fraction of occupied nodes in the network. However, in strongly interdependent networks, the transition is discontinuous "jumping" from 0 to ≈ 0.4 at the percolation threshold. In other words, the giant component suddenly forms in this case, where in weakly interdependent networks it forms gradually. (b) n fully interdependent networks (all nodes mutually depend on nodes in other networks). As the number of networks n increases, the "jump" in the size of the largest connected component is bigger. Figure adapted from [103].

As stated earlier, the model of Buldyrev *et al.* has attracted considerable attention and consequently has been extended in various ways [103]. For example, in a system of two interdependent ER networks where only part of the nodes are interdependent (called partially

instead of fully interdependent networks), there exists a critical fraction of interdependent nodes above which the system is very fragile, exhibiting a first order percolation transition [180]. In case the networks are scale-free, a very interesting "hybrid" phase between first and second order was found, where the size of the giant component has a sharp drop from finite value to a much smaller, yet a nonzero value [249]. This "hybrid" transition was also found in a system of two ER networks interconnected by both connectivity and dependency links [127].

A theoretical framework for understanding the robustness of interdependent networks under targeted attacks on specific degree nodes was developed in [130], and it was found that even when hubs have low failure probability (i.e. nodes are attacked in order from lowest degree to highest degree), interdependent scale-free networks have percolation threshold much larger than zero. This is because high-degree nodes often depend on the more common low-degree nodes in the other network. And indeed, interdependent networks where nodes depend on similar-degree nodes (i.e. when the inter-degrees are positively correlated), called intersimilar networks, are much more robust [54, 128, 179]. However, degree correlation within a single network (i.e., the likelihood of nodes with similar degree to be connected), as well as high clustering coefficient, has been shown to decrease the robustness of the entire system [131, 206, 239, 250]. Recently, there has been several efforts to extend the theory of interdependent networks from two networks to a network of interdependent networks (NON) [40, 41, 101, 102]. Also, recent papers examine design strategies for mitigating risk in systems of interdependent networks, which can be done at a very low cost with careful selection of autonomous nodes (nodes which are not interdependent) [202, 229].

Finally, since one of the main motivation for the work of Buldyrev *et al.* was to address catastrophic cascade of events in critical infrastructures, and these are often spatially embedded (e.g. power grids, Internet), several models of interdependent networks were developed taking spatial constraints into account. For example, it has been shown that in a system composed of two interdependent square lattice networks placed on the same Cartesian plane, there exist a critical length of dependency links (i.e. the maximum distance between a node in one network and the node that it depends on in another network), r_c, above which the system is very fragile, characterized by an abrupt collapse of the giant component (first order transition) [75, 148]. In a system of n interdependent networks, r_c has been shown to rapidly decrease with n [209]. This result was also verified in a system of two identical random regular graphs (instead of lattices), where networks with long dependency distances (shortest path between a pair of interdependent nodes) are much more vulnerable than networks with short dependency distances [140]. A very strong result was obtained in [30], where interdependent spatially embedded networks with randomly assigned interdependency links (i.e. with no length constraint) were found to be

extremely vulnerable: the failure of *any* small fraction of interdependent nodes leads to an abrupt collapse. Finally, Berezin *et al.* [37] studied localised geographical attacks in two interdependent spatially embedded networks and found a critical attack size above which it will spread through the entire system and lead to its complete collapse.

2.4.2 Interacting, modular, and multiplex networks

The exciting discoveries made in the burst of studies of interdependent networks described above, have raised the need to consider other scenarios where interactions between networks, and between networks and their environment occur. One of the most useful models, called *interacting networks* or *interconnected networks* is considering a set of networks in which some of the nodes from the various networks are adjacent to each other (also called *coupled* nodes) [85, 146]. But unlike interdependent networks, intra-network edges (connecting nodes within the same network) and inter-network edges (connecting nodes from different networks) are semantically equal, both representing connectivity relations just like in a single network. Examples for such networks include interconnected power grids owned by different utilities [50], coupled climate networks represent different isobaric surfaces [85], and interacting transportation networks such as an underground and a railroad network [223] or an airport and port network [179].

A special case of interacting networks, called *multiplex networks*, is when the node sets are the same across the different networks, representing the case where a set of elementary units are connected by relationships of different kinds [34, 39]. Examples include banks connected through various credit relations [24], airports connected through different commercial airlines [58] and YouTube users connected through various types of sharing activities (e.g. shared subscriptions, shared favourite video) [76].

Both multiplex and interacting networks are not fundamentally different than single networks, and could be mathematically treated as ones. A multiplex network, for example, could be studied by examining one layer at a time or by aggregating layers. Similarly, interacting networks could be simply thought as subgraphs of the same network. However, modelling systems as coupled networks has been shown to be extremely useful in some cases. For example, treating an heterogeneous system as a multiple interacting networks allows its analytical study, which might otherwise be impossible. This observation is not entirely new and the need to analyse heterogeneous systems was raised in previous studies (before the "interdependent networks era"), mainly in the context of *modular networks* [15, 86, 106, 155, 172, 213, 231]. Modular networks are consisting of tightly-connected groups of nodes (modules) with less connections between the groups, which often results in heterogeneous degree distributions [86, 106], degree correlation (the tendency of similar-degree nodes to be connected) [155, 172, 213, 231], and occupation

probability among modules [15]. Similarly, in cases where the behaviour of nodes is different among subnetworks, such as power grids owned by different utilities with different policies, the separation into networks can be very useful [50]. Multiplex networks are also not a new concept. In social systems for example, the need categorise edges based on the nature of the relationships was already expressed in the 70s [47, 233, 242].

But although some aspects of coupled networks were occasionally studied before, it is only recently that network scientists are making extensive and collective efforts to develop analytical frameworks to study coupled complex systems [139]. And indeed, as we demonstrate in this section, the structure and function of coupled networks can not be explained by reducing them to single networks. The new degrees of freedom that result from the introduction of multiple layers often give rise to completely"new physics" than in single networks [139].

And just like the network science "revolution" that took place in the late 90s with the availability of large data sets, which were found to be very different than random graphs, the increasing availability of empirical data for fundamentally multilevel systems is one of the main driving forces behind the growing interests in coupled networks. Considering the interplay and the interactions between the different layers in a system, can often reveal a lot of valuable information about its structure and evolution. For example, the structural properties of the European Air Transportation Network are generally not present in single layers, correspond to commercial airlines, rather they are the consequence of an emerging phenomenon intimately related to the multilayer character of the system [58]. In the international trade network, commodity-specific trade relations among world countries were found to exhibit a very rich and different structure, suggesting that the multinetwork is emerging as a nontrivial aggregation of several interdependent layers [25]. By separating climate networks, where geographical locations are connected via correlation relationship, into different isobaric surfaces, Donges *et al.* identified regions which are particularly important for mediating vertical wind field interactions [85]. In chapter 6, we apply a multilayer network approach to the study of underground and street networks in the metropolitan areas of New York and London, revealing important structural features that arise from the interplay between transportation systems in urban areas.

Once realising that the structure of coupled multilevel systems is a result of nontrivial interplay between its layers, it is pressing to understand how this complex structure affects dynamical processes taking place on coupled networks. One of the first attempt to develop new analytical tools that describe the behaviour of coupled networks was made by Leicht and D'Souza [146]. They introduced an analytical framework, based on generating functions, to study percolation process in interacting networks, and showed that the percolation threshold in an individual network can be significantly lowered once interconnections to other networks are considered,

meaning that interconnections could enhance the robustness of interacting networks. We have later extended their framework to study the resilience of modular networks to international attacks, revealing an interesting behaviour characterised by two percolation regimes, see chapter 3. Brummitt *et al.* studied cascades of load shedding in interacting power grids and found that coupling could mitigate large avalanches in the individual networks, as the coupled networks provide reservoirs to absorb excess load. However, yet when compared to isolated networks, interconnected networks more frequently suffer avalanches that are large in both networks [50]. This result was also confirmed by Tan *et al.* [227], who studied cascading failures of loads in interconnected networks under intentional attack and found that for sparse coupling, enhancing the coupling probability can make interconnected networks more robust against intentional attacks, but to keep increasing the coupling probability has the opposite effect for dense coupling.

Percolation processes were also extended to multiplex networks, and it was suggested that interlayer degree correlations (where nodes who have high degree in one layer are likely to have high degree in another layer) increase the robustness of multiplexes to random failure [64, 145, 159]. Motivated by the striking difference in the interlayer degree correlations that were found in biological networks compared to other networks [64, 179], we later showed that this result might not hold in certain situations, providing a possible explanation to the uncorrelated randomized structure found in protein networks, see chapter 5. Finally, Watts cascade dynamics, where nodes are activated if a fraction of their active nodes is exceeding a threshold, was studied on a multiplex in [51], showing that layers unsusceptible to global cascades can cooperatively achieve them if coupled, meaning that multiplex networks are generically more vulnerable to global cascades than simplex networks.

The extension of spreading processes to coupled networks has been shown to yield reach and complex dynamics that can not be obtained in single networks. For example, unlike in single networks where the epidemic either breaks or dies out, in interacting networks, a mixed phase can emerge, where an epidemic occurs in one network but does not spread to the coupled network [82]. We have shown that in case of adaptive networks, where the epidemic spreading is coupled to a rewiring process between susceptible and infected nodes, interconnections can result in a localise epidemic breakout only in the coupled nodes, see chapter 4. In another study, it was shown that a global endemic state may arise in the coupled system even though the epidemics is not able to propagate on each network separately and even when the number of coupling connections is small [198]. These studies have important implications for disease prevention policies, for example, by determining the effectiveness of travel restrictions, which can be modelled as interconnection between geographically distance populations. Multiplex networks can also be used to study the interrelation between the spreading of an epidemics, and

the information awareness to prevent its infection [99, 110].

Other types of diffusion processes were studied on coupled networks, further illustrating the physical implications of multilayer structure. For example, a super-diffusive behavior was found in multiplex networks, where the time scales of diffusion associated to the multiplex are shorter than in any particular individual layer network [107, 216]. Morris *et al.* [165] studied the utility of coupled spatial networks and found two regimes depending on the distribution of sources and sinks between which a flow process is taking place. When the flows go from many sources to a small number of sinks, the network utility is largest when the coupling is at its maximum. However, when many sources correspond to many sinks, the optimal coupling is no longer maximal as flow congestions must be taken into account. Link overlap (where links between two nodes exist in more than one layer or network) and inter- and intra-degree correlation (where the degrees of a node and its coupled node and or the degrees of the same node in different layers are correlated), were found to have dramatic dynamical consequence [191, 192, 203, 226]. Halu *et al.* showed that spatial multiplex ensembles naturally develop a significant overlap of the links, and demonstrate this finding in an interacting airport and railway networks in India [120]. Finally, the coupling of multiple networks have been shown to promote cooperation in evolutionary games, such as the Prisoner's Dilemma [108, 109], and to provide stabilisation and synchronization mechanism, even when individual layers work on the chaotic regime, raising the need for a conceptual transition from the physics of single-layered networks to the physics of coupled networks [71, 217].

2.5 Conclusions

In the last 15 years, network science, the study of the structure and dynamics of complex networks, has rapidly evolved and established itself as a leading scientific field in the description of complex systems. Understanding networks of interactions has been repeatedly shown as the key to understanding emergent collective phenomena in the complex systems they represent. By looking at real data, network scientists discovered that real networks have very different connectivity patterns from their homogeneous random counterparts that were widely used before, and these complex nontrivial patterns have tremendous dynamical implications, such as the lack of percolation threshold in scale-free networks.

In today's world of big-data, it has become possible to empirically observe the interactions, growth and evolution of networks. Existing network science tools are no longer sufficient to describe the phenomenon we observe. Large-scale electrical blackouts, resulting from tight interdependencies between infrastructures, can not be explained using the traditional single

isolated network models. In an attempt to tackle this challenge, in the past four years we are observing a paradigm shift in the study of networks. Coupled networks are the next generation network models, used to describe and study the behaviour of coupled multilevel complex systems, common in nature and engineering. This new type of networks has been shown to give rise to completely new phenomena, thus attracting so many network scientists, including myself, to contribute to this significant transition in the history of network science.

By design, coupled networks contain more information than single isolated networks, and are bound to give a more detailed description of the systems they represent. But where for some applications, there might be an obvious way to identify the subnetworks or layers consisting a system, in others, one could face a challenging model selection task. This is one of the problems typical to such a young and active field, where researchers are still exploring the new and exciting opportunities it offers, in a combining effort to understand the complex systems all around us. This thesis is hopefully a valuable contribution to this collective effort.

CHAPTER THREE

RESILIENCE OF MODULAR NETWORKS

Many real networks including social, technological, and biological networks have been shown to exhibit a modular structure, where a number of tightly-connected groups of nodes (modules) have relatively few interconnections. While empirical and numerical studies have repeatedly demonstrated the importance of modularity structure for the functionality of systems, comprehensive theoretical frameworks explaining its effects are still lacking. Based on recent breakthroughs in the understanding of coupled multilayer networks, we are able to systematically study modularity structure, and its effect on the vulnerability of networks. In this chapter, we study analytically and numerically the resilience of modular networks to attacks on interconnected nodes, those connecting between the modules, which are often more exposed to failure. The chapter is based on the author's work [205].

3.1 Introduction

Modularity, also called community structure, corresponds to the multipartite organisation of large-scale systems consisting of cohesive groups on nodes, called modules of communities, see Fig. 3.1. The are several different ways in which the structural cohesion of nodes can be quantified, for example, Newman's modularity [173] measures the number of links falling within modules (intralinks) minus the expected number in an equivalent network with links placed at random. But the important point here is that the heterogeneous division of nodes into modules, or into subnetworks in a system of interacting networks, is playing a key role in the global behaviour of the system. The modular organisation of the Internet, and other large-scale infrastructures, tremendously enhances scalability and diffusion processes [91, 118]. Modules

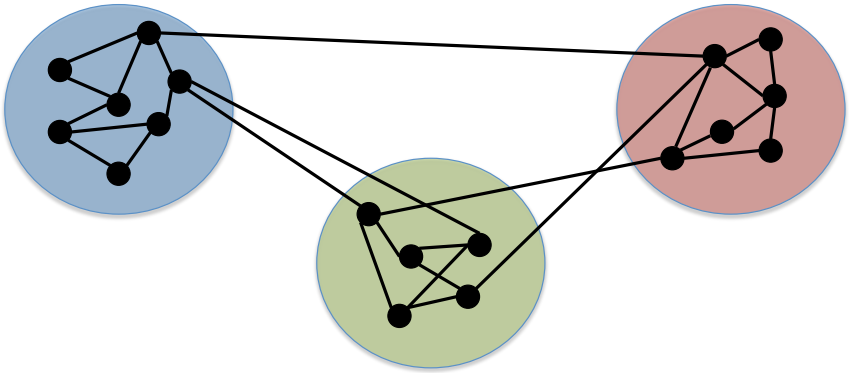


Figure 3.1: A schematic representation of a modular network consisting of three groups of nodes (modules) with dense internal connections and sparser connections between groups.

of protein complexes and dynamic functional units constitute the building blocks of molecular networks [218]. Individuals are divided into social or geographical regions with strong local ties promoting economic development and knowledge creation [38, 104]. Finally, the non-random modular architecture of neural networks is considered crucial for the brain's functional demands of segregation and integration of information [31, 56], and disrupted brain modular organisation is related to neuropathology, such as schizophrenia [14], autism [29], Alzheimer [230] and impulsivity [77].

Although most research so far has focused on the detection of modularity structure [96], a few studies have examined its affect on the functionality and dynamics of networks. For example, Wu *et al.* [245] studied numerically cascade propagation in modular networks, where heavy loaded nodes fail and redistribute their load to other nodes, and found that networks with a distinct partition into modules are more robust to such dynamics. Some analytical approaches were developed to study dynamics on networks with assortativity structure, which divides the network into groups of nodes with similar properties [106, 155, 172]. They found that networks that are assortatively mixed by degree (i.e. divided into modules of similar degree nodes) are more robust to both random failure and targeted attack of the highest-degree nodes, although recently it has been shown that the case is different in interdependent networks [250]. Finally, Bagrow *et al.* [18] have studied a percolation process on a network of modules, where a module becomes nonfunctional if a critical fraction of its nodes fail, leading to the isolation of modules long before the network itself falls apart.

But despite these recent advances, many questions regarding the implications of modularity

structure to networks dynamics still remain. In particular, empirical and numerical studies of brain and interactome networks suggest that the deletion of nodes connecting between modules can have a deleterious effect on the network integrity [121], efficiency [220], and stability [123]. However, an analytical framework for understanding this observed phenomena and predicting the vulnerability of these systems is still lacking. Through recent advances in the understanding of coupled multilayer networks [53, 78, 103, 139, 146] however, it is now possible to study systematically both modularity and its effects on network vulnerability.

In this chapter, we develop an analytical framework for studying the robustness of modular networks in the presence of attack on interconnected nodes, those connecting between the modules, which are often more exposed to failure. We study a percolation process on networks consisting of a varying number of modules, m, and a varying number of interconnected nodes. The analytical solution reveals two percolation regimes separated by a critical number of modules m^*: for $m < m^*$ one needs to remove all interconnected nodes to break the system, while the modules are almost unaffected internally. In contrast, for $m > m^*$ one needs to remove only a fraction of the interconnected nodes, before the system collapses. This is due to the fact that for $m > m^*$ the number of interconnected nodes is high and partial removal of these breaks the modules internally, which helps to bring about the rapid collapse of the whole system. In other words, our analytical formalism provides the critical concentration of interconnections between modules, above which the internal structure of each module is inseparable from the system as a whole. Our approach can also be used to study analytically attacks on high betweenness centrality nodes, which in modular structures, correspond to interconnected nodes. Such attacks, which have only been studied numerically so far, are considered to be among the most harmful attack strategies [125, 201].

The rest of this chapter is organised as follows: in section 3.2 we present a model for generating random modular networks with a varying number of modules. In section 3.3 we present an analytical formalism based on generating functions for studying percolation processes in these networks. In section 3.4 we discuss the analytical predictions and verify them with extensive computer simulation, and finally, in section 3.5 we summarise our findings and discuss their implications.

3.2 Model of random modular networks

We consider a random modular network with N nodes divided into m equal sized modules. Similarly to [174], we define p_{intra} as the probability to connect nodes in the same module and p_{inter} as the probability to connect nodes in different modules. Thus, the total number of

intra-module (inter-module) links is given by the probability for a link p_{intra} (p_{inter}) multiplied by the number of possible links yielding

$$M_{\text{intra}} = p_{\text{intra}} \frac{N(\frac{N}{m}-1)}{2}, \tag{3.1}$$

$$M_{\text{inter}} = p_{\text{inter}} \frac{N(m-1)\frac{N}{m}}{2}. \tag{3.2}$$

We define α to be the ratio between the probabilities for an intra- and inter-module link

$$\alpha = \frac{p_{\text{intra}}}{p_{\text{inter}}}. \tag{3.3}$$

Note that the ratio between the number of inter-modules links and intra-module links depends not only on α, but also on the number of modules

$$\frac{M_{\text{intra}}}{M_{\text{inter}}} = \frac{p_{\text{intra}}(\frac{N}{m}-1)}{p_{\text{inter}}(m-1)\frac{N}{m}} \sim \frac{p_{\text{intra}}}{p_{\text{inter}}(m-1)} = \frac{\alpha}{m-1} \tag{3.4}$$

with the approximation correct for large modules size $\frac{N}{m} \to \infty$. Thus, our model is taking into consideration that systems comprised of more modules have more inter-links accordingly, as illustrated in Fig. 3.2.

The model described above can be used to produce any random modular network with a given number of nodes, links, modules and a ratio between the density of intra- and inter-links. It thus provides, similar to other network formation models, an alternative for having a network with modular structure given to us up-front from which we can then extract the topology. In this chapter we focus on modular Erdős-Rényi (ER) networks where both the intra-module and inter-module connectivity are Poisson distributed with means k_{intra} and k_{inter} respectively. At end of the chapter we also show numerical results for modular scale-free networks. Given N, M, m and α, we generate a modular ER network in the following way

- **Compute the number of inter- and inter-module links :** we extract M_{intra} and M_{inter} from Eq. 3.4 and the fact that $M_{\text{intra}} + M_{\text{inter}} = M$.
- **Build each module internally :** we build m ER networks, each with $\frac{N}{m}$ nodes and $\frac{M_{\text{intra}}}{m}$ links, using the networkX graph generator method fast_gnp_random_graph, which implements an algorithm for efficient generation of large sparse random networks [33].
- **Connect between the modules :** we randomly choose M_{inter} couples of nodes from two different modules, and connect them.

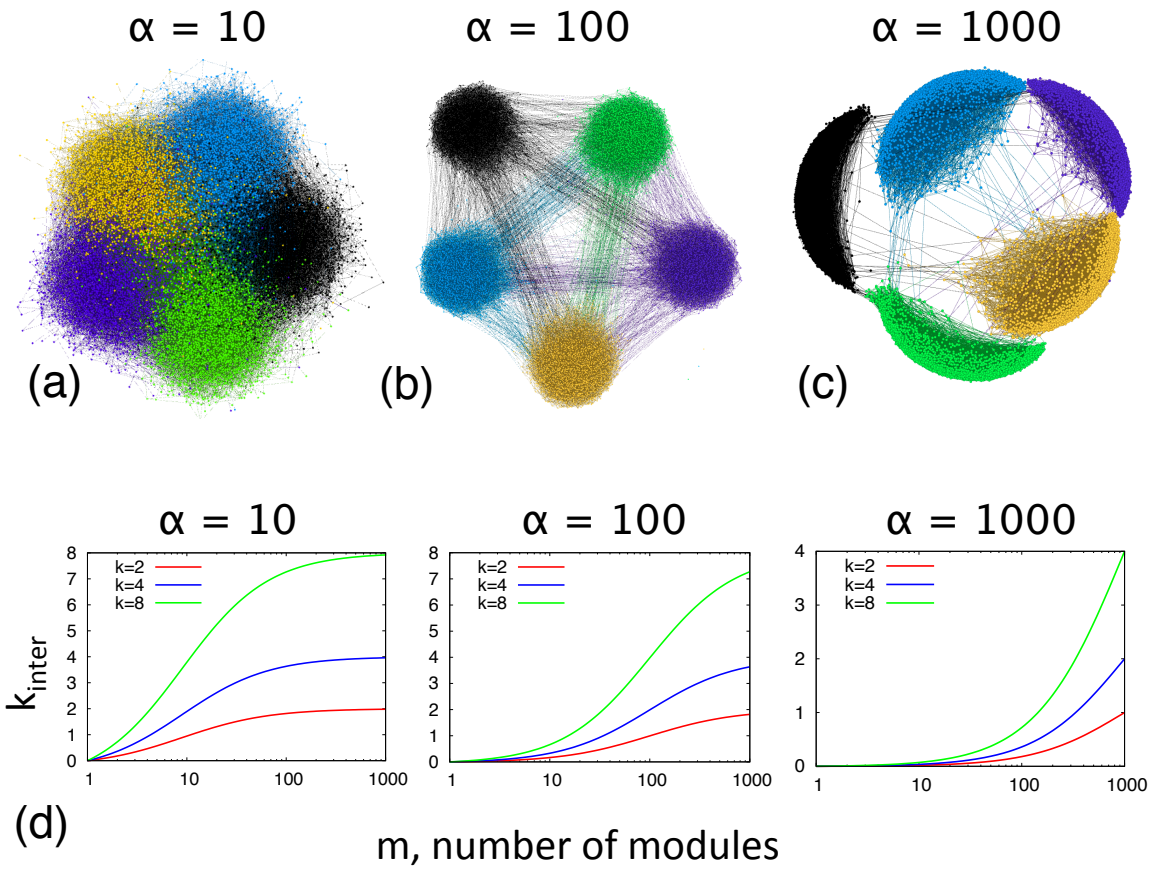


Figure 3.2: (a)-(c) Illustration of the effect of α on the obtained modular network using Gephi with force atlas layout [32, 178], on a network of size $N = 10000$ with mean degree $k = 8$ divided into $m = 5$ modules. (d) Average inter-module degree, k_{inter}, as a function of the number of modules, m, for different mean degrees, k, and α, calculated from Eq. 3.2.

In Fig. 3.2(a)-(c) we present an example of modular networks generated with different values of α, and visualised using force-directed layout, which has been shown to demonstrate network modularity [178]. To illustrate the effect of the number of modules on the number of interconnected nodes (see Eq. 3.4), in Fig. 3.2(d) we show the increase of the mean inter-degree, k_{inter}, as a function of m. As m goes to infinity, the mean inter-degree is converging to the total mean degree of the network $k = k_{\text{intra}} + k_{\text{inter}}$ at a rate affected by α.

3.3 Analytical framework

In the following we study the percolation properties of the model described above. First, we describe the formalism developed [146] to study a percolation process in systems comprised of multiple interconnected networks (i.e. network of networks), and we provide a full analytical

solution for modular ER networks. Then, using the extension of the generating functions approach to general attacks developed by Callaway *et al.* [57], we study the percolation properties of modular networks in case of attack on interconnected nodes.

3.3.1 Generating functions for networks of networks

We begin by describing the formalism developed by Leicht *et al.* [146], which extends the generating function approach for single networks (described in details in section 2.3.1) to interacting and modular networks, which are essentially the same mathematical entity. Consider a system with $m \geq 2$ interacting networks (which can also be considered as modules), described by a set of degree distributions. Each individual network is characterized by a multi-degree distribution, $\{p^i_{k_1 k_2 \ldots k_m}\}$, where $p^i_{k_1 k_2 \ldots k_m}$ is the fraction of all nodes in the network i that have k_1 links to nodes in network 1, k_2 links to nodes in network 2, etc. The multi-degree distribution for each network may be written in the form of a generating function

$$G_i(x_1 x_2 \ldots x_m) = \sum_{k_1 k_2 \ldots k_m = 0}^{\infty} p^i_{k_1 k_2 \ldots k_m} {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m}. \tag{3.5}$$

Note that this function is simply an extension of the single-network generating function defined in Eq. 2.6. The partition of a node's degree into m degrees, corresponding to its number of connections in each subnetwork, allows a finer analysis of heterogeneous systems, which is essentially the objective of studying modular or interacting networks.

Similarly to the excess degree in a single network defined in section 2.3.1, we define the multi excess degree, $q^{ij}_{k_1 \ldots k_j \ldots k_m}$ as the probability of following a randomly chosen ij-link (i.e. a link falling between a node in network i and a node in network j) and arrive to an i-node (node in network i) with additional k_j links apart from the one by which it has been reached (also called outgoing links), to nodes in network j. This probability is proportional to k_j (since the i-node has k_j possible links through which we can arrive), and therefore the associated generating function is given by

$$\begin{aligned} G_{ij}(x_1 x_2 \ldots, x_m) &= \sum_{k_1 k_2 \ldots k_m = 0}^{\infty} q^{ij}_{k_1 k_2 \ldots, k_m} {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m} \\ &= \frac{\sum_{k_1 k_2 \ldots k_m = 0}^{\infty} (k_j + 1) p^i_{k_1 \ldots (k_j+1) \ldots k_m} {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m}}{\sum_{l_1 l_2 \ldots l_m = 0}^{\infty} (l_j + 1) p^i_{l_1 \ldots (l_j+1) \ldots l_m}} \end{aligned}$$

$$= \Big(\frac{\partial}{\partial x_j} \underbrace{\sum_{l_1 l_2 \ldots l_m=0}^{\infty} p^i_{l_1 l_2 \ldots l_m}}_{G_i(\mathbf{1})}\Big)^{-1} \frac{\partial}{\partial x_j} \underbrace{\sum_{k_1 k_2 \ldots k_m=0}^{\infty} p^i_{k_1 k_2 \ldots k_m} {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m}}_{G_i(\mathbf{x})}$$

$$= \Big(\frac{\partial G_i}{\partial x_j}(\mathbf{1})\Big)^{-1} \frac{\partial G_i}{\partial x_j}(\mathbf{x}) \tag{3.6}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ and $\mathbf{1} = (1, 1, \ldots, 1)$. Note that the second transition from sum to partial derivative is a "shift left" of the function, just as the one done in Eq. 2.10.

Let us now consider the distribution of the sizes of components reachable by following a randomly chosen ij-link to a i-node and then following its additional outgoing links, and let $H_{ij}(\mathbf{x})$ be the associated generating function. Summing over all the types of connectivity possible for the i-node leads to the self consistency equation (see [146] for full details)

$$\begin{aligned} H_{ij}(\mathbf{x}) &= x_i \sum_{k_1 k_2 \ldots k_m=0}^{\infty} q^{ij}_{k_1 k_2 \ldots, k_m} H_{1i}(\mathbf{x})^{k_1} \ldots H_{mi}(\mathbf{x})^{k_m} \\ &= x_i G_{ij}(H_{1i}(\mathbf{x}), H_{2i}(\mathbf{x}), \ldots, H_{mi}(\mathbf{x})). \end{aligned} \tag{3.7}$$

Starting from a randomly chosen i-node, rather than a random ij-link, the distribution of the sizes of components reachable by following its additional outgoing links is generated by

$$H_i(\mathbf{x}) = x_i G_i[H_{1i}, H_{2i}, \ldots, H_{mi}]. \tag{3.8}$$

Using the generating function $H_i(\mathbf{x})$, we can now calculate the average number of j-nodes in the component of a randomly chosen i-node

$$\langle s_i \rangle_j = \frac{\partial H_i}{\partial x_j}(\mathbf{x})|_{\mathbf{x}=\mathbf{1}} = \delta_{ij} + \sum_{l=1}^{m} \frac{\partial G_i}{\partial x_l}(\mathbf{1}) \frac{\partial H_{li}}{\partial x_j}(\mathbf{1}) \tag{3.9}$$

where δ_{ij} denotes the Kronecker delta.

3.3.2 Solution for modular ER networks

We proceed now to find the percolation threshold (the critical occupation probability above which the giant component emerges) for modular ER networks, consisting of m modules, as described before. Assuming that the intra-module and inter-module degrees are uncorrelated, the

joint degree distribution of the network is given by

$$p^{i}{}_{k_1,k_2,\ldots,k_m} = \frac{{k_{\text{intra}}}^{k_i} e^{-k_{\text{intra}}}}{k_i!} \prod_{j \neq i} \frac{{\frac{k_{\text{inter}}}{m-1}}^{k_j} e^{-\frac{k_{\text{inter}}}{m-1}}}{k_j!}. \tag{3.10}$$

By the Taylor series expansion for exponential function, we obtain that the generating function of a Poisson degree distribution is with mean z is $e^{z(x-1)}$, and similarly in our case

$$G_i(\mathbf{x}) = e^{k_{\text{intra}}(x_i-1)} e^{\sum_{j \neq i} \frac{k_{\text{inter}}}{m-1}(x_j-1)}. \tag{3.11}$$

From Eq. 3.6, the generating function for the excess degree distribution is given by

$$G_{ij}(\mathbf{x}) = \begin{cases} \frac{\frac{\partial G_i}{\partial x_j}(\mathbf{x})}{\frac{\partial G_i}{\partial x_j}(\mathbf{1})} = \frac{k_{\text{inter}}}{m-1} G_i(\mathbf{x}) \frac{m-1}{k_{\text{inter}}} = G_i(\mathbf{x}), & \text{if } j \neq i \\ \frac{\frac{\partial G_i}{\partial x_j}(\mathbf{x})}{\frac{\partial G_i}{\partial x_j}(\mathbf{1})} = k_{\text{intra}} G_i(\mathbf{x}) \frac{1}{k_{\text{intra}}} = G_i(\mathbf{x}), & \text{otherwise} \end{cases} \tag{3.12}$$

yielding $G_{ij}(\mathbf{x}) = G_i(\mathbf{x})$ for all j. Thus, we obtain $H_{ij}(\mathbf{x}) = H_i(\mathbf{x})$, and then equation 3.9 can be written as

$$\langle s_i \rangle_j = \delta_{ij} + k_{\text{intra}} \frac{\partial H_i}{\partial x_j}(\mathbf{1}) + \frac{k_{\text{inter}}}{m-1} \sum_{\substack{l=1 \\ l \neq i}}^{m} \frac{\partial H_l}{\partial x_j}(\mathbf{1}) \tag{3.13}$$

where

$$\frac{\partial H_i}{\partial x_j}(\mathbf{1}) = \delta_{ij} + k_{\text{intra}} \frac{\partial H_i}{\partial x_j}(\mathbf{1}) + \frac{k_{\text{inter}}}{m-1} \sum_{\substack{l=1 \\ l \neq i}}^{m} \frac{\partial H_l}{\partial x_j}(\mathbf{1}). \tag{3.14}$$

For example, using the notation $h_i = \frac{\partial H_i}{\partial x_1}(\mathbf{1})$, the system obtained for $\langle s_1 \rangle_1$ is:

$$\begin{aligned} h_1 &= 1 + k_{\text{intra}} h_1 + \frac{k_{\text{inter}}}{m-1}(h_2 + h_3 + \cdots + h_m) \\ h_2 &= k_{\text{intra}} h_2 + \frac{k_{\text{inter}}}{m-1}(h_1 + h_3 + \cdots + h_m) \\ &\vdots \\ h_m &= k_{\text{intra}} h_m + \frac{k_{\text{inter}}}{m-1}(h_1 + h_2 + \cdots + h_{m-1}) \end{aligned} \tag{3.15}$$

Summing equations for $h_2, \dots h_m$, we obtain:

$$\begin{aligned} h_2 + h_3 + \ldots + \ldots h_m &= k_{\text{inter}} h_1 + (k_{\text{intra}} + \frac{m-2}{m-1} k_{\text{inter}})(h_2 + h_3 + \cdots h_m) \\ \Rightarrow h_2 + h_3 + \cdots h_m &= \frac{k_{\text{inter}} h_1}{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}} \end{aligned} \tag{3.16}$$

Substituting into 3.15, we obtain:

$$\begin{aligned} h_1 &= 1 + k_{\text{intra}} h_1 + \frac{k_{\text{inter}}}{m-1}(h_2 + h_3 + \cdots + h_m) \\ &= 1 + k_{\text{intra}} h_1 + \frac{k_{\text{inter}}}{m-1}\left(\frac{k_{\text{inter}} h_1}{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}}\right) \\ &\Rightarrow h_1\left(1 - k_{\text{intra}} - \frac{\frac{k_{\text{inter}}^2}{m-1}}{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}}\right) = 1 \\ &\Rightarrow h_1 = \frac{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}}{(1 - k_{\text{intra}})(1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}) - \frac{{k_{\text{inter}}}^2}{m-1}}. \end{aligned} \tag{3.17}$$

h_1 diverges when $(1-k_{\text{intra}})(1-k_{\text{intra}} - \frac{(m-2)k_{\text{inter}}}{m-1}) - \frac{{k_{\text{inter}}}^2}{m-1} = 0$, which yields $k = k_{\text{intra}} + k_{\text{inter}} = 1$ for every m. This is also where all h_i diverge, marking the point where the giant component emerges, recovering the standard result for single networks without modules (see section 2.3.1). Thus, in the case of random node failure the percolation threshold only depends on the mean degree, k. This result has been shown in [146] for a system of two networks and here we provided the equation for general m.

3.3.3 Attack on interconnected nodes

But in contrast to random node failure, in real systems the interconnected nodes are often more exposed to failure than other nodes. For example, it has been shown that ageing and schizophrenia could result in a damage to the interconnected nodes in brain networks [156, 210]. In addition, it is often the case where interconnected nodes are considered to be important, such as airport hubs like the New York City and London airports, which provide an attractive target for attacks [118]. Finally, as we mentioned in the introduction, empirical studies of biological networks suggest that these are very susceptible to the removal of interconnected nodes. Therefore, in the following, we consider an attack on modular ER networks where the interconnected nodes are randomly removed.

We extend Callaway *et al.*'s approach [57] for studying the robustness of networks to intentional

attacks, from single to multiple networks in a similar manner that was done by Leicht *et al.* [146]. In this approach the occupation probability of nodes is not constant as before, but is a function of the node's degree (see also section 2.3.1). Consider a node from module i with k_1 links in module 1, k_2 links in module 2 and etc, and let $r_i(k_1,k_2,\ldots,k_m)$ denote its occupation probability. When interconnected nodes (i.e. nodes with at least one interconnection) are randomly removed, this probability is given by

$$r_i(k_1,k_2,\ldots,k_m) = \begin{cases} 1, & \text{if } \sum_{\substack{j=1 \\ j\neq i}}^{m} k_j = 0 \\ q, & \text{otherwise} \end{cases}, \tag{3.18}$$

where q is the probability that a randomly chosen interconnected node is occupied, i.e. the fraction of occupied interconnected nodes. In other words, nodes with no interconnections are never removed (i.e. occupied with probability 1) while interconnected nodes are removed with probability $1-q$. Let p be the general occupation probability, i.e. the probability that a randomly chosen node is occupied. Since the probability for a node to be interconnected is $1-e^{-k_{\text{inter}}}$ (one minus a Poisson distribution with mean k_{inter} at 0), we obtain

$$q = \frac{p - e^{-k_{\text{inter}}}}{1 - e^{-k_{\text{inter}}}}. \tag{3.19}$$

We define the generating functions for the degree and excess degree distributions of occupied nodes (which is, once again, an extension of Eq. 2.19 for a single network)

$$F_i(x_1,x_2,\ldots,x_m) = \sum_{k_1,k_2,\ldots,k_m=0}^{\infty} p^i_{k_1,k_2,\ldots,k_m}\, r^i_{k_1,k_2,\ldots,k_m}\, {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m} \tag{3.20}$$

$$F_{ij}(x_1,x_2,\ldots,x_m) = \sum_{k_1,k_2,\ldots,k_m=0}^{\infty} q^{ij}_{k_1,k_2,\ldots,k_m}\, r^i_{k_1,k_2,\ldots,k_m}\, {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m}. \tag{3.21}$$

By substituting the occupation probability defined in 3.19 into equation 3.20 we obtain

$$\begin{aligned} F_i(\mathbf{x}) &= \underbrace{\sum_{k_i=0}^{\infty} p^i_{0\ldots k_i \ldots 0}\, {x_i}^{k_i}}_{\text{inter-degree is zero}} + \underbrace{\sum_{\substack{k_1,k_2,\ldots,k_m=0 \\ (k_1,\ldots,k_{i-1},k_{i+1},\ldots,k_m)\neq 0}} p^i_{k_1 k_2 \ldots k_m}\, {x_1}^{k_1} {x_2}^{k_2} \ldots {x_m}^{k_m}}_{\text{inter-degree is not zero}} \\ &= e^{-k_{\text{inter}}} e^{k_{\text{intra}}(x_i-1)} + q\left(G_i(\mathbf{x}) - e^{-k_{\text{inter}}} e^{k_{\text{intra}}(x_i-1)}\right) \\ &= e^{k_{\text{intra}}(x_i-1)-k_{\text{inter}}}(1-q) + qG_i(\mathbf{x}). \end{aligned} \tag{3.22}$$

By substituting 3.22 into equation 3.21, we obtain

$$F_{ij}(\mathbf{x}) = \begin{cases} \frac{\partial F_i}{\partial x_j}(\mathbf{x})\frac{m-1}{k_{\text{inter}}} = qG_i(\mathbf{x}), & \text{if } i \neq j \\ \\ \frac{\partial F_i}{\partial x_i}(\mathbf{x})\frac{1}{k_{\text{intra}}} = F_i(\mathbf{x}), & \text{otherwise.} \end{cases} \tag{3.23}$$

Let $J_{ij}(\mathbf{x})$ be the generating function for the distribution of the number of occupied nodes in the component reachable by following a randomly chosen ij-link to a i-node and then following its additional outgoing links. This component contains zero nodes if the node at the end of the link in question is unoccupied, which happens with probability $1 - F_{ij}(\mathbf{1})$, or the link may lead to an occupied i-node with k_j other links leading out of it to nodes in module j, distributed according to $F_{ij}(\mathbf{x})$. Thus $J_{ij}(\mathbf{x})$ satisfies the self-consistency condition:

$$J_{ij}(\mathbf{x}) = 1 - F_{ij}(\mathbf{1}) + x_i F_{ij}[J_{1i}, J_{2i}, \ldots, J_{mi}]. \tag{3.24}$$

And similarly, the distribution of the number of nodes reachable from a randomly chosen i-node is generated by:

$$J_i(\mathbf{x}) = 1 - F_i(\mathbf{1}) + x_i F_i[J_{1i}, J_{2i}, \ldots, J_{mi}] \tag{3.25}$$

Using $J_i(\mathbf{x})$ we can now calculate the average number of occupant j-nodes in the component of a randomly chosen i-node,

$$\begin{aligned} \langle s_i \rangle_j &= \frac{\partial J_i}{\partial x_j}(\mathbf{x})|_{\mathbf{x}=1} \\ &= \delta_{ij} F_i(\mathbf{1}) + \frac{\partial F_i}{\partial x_j}(\mathbf{1})\frac{\partial J_{ii}}{\partial x_j}(\mathbf{1}) + \sum_{\substack{l=1 \\ l\neq i}}^{m} \frac{\partial F_i}{\partial x_l}(\mathbf{1})\frac{\partial J_{li}}{\partial x_j}(\mathbf{1}) \\ &= \delta_{ij} F_i(\mathbf{1}) + k_{\text{intra}} F_i(\mathbf{1})\frac{\partial J_{ii}}{\partial x_j}(\mathbf{1}) + q\frac{k_{\text{inter}}}{m-1}\sum_{\substack{l=1 \\ l\neq i}}^{m} \frac{\partial J_{li}}{\partial x_j}(\mathbf{1}) \end{aligned} \tag{3.26}$$

For example, using the notation $j_{st} = \frac{\partial J_{st}}{\partial x_1}(\mathbf{1})$, the system obtained for $\langle s_1 \rangle_1$ is:

$$\begin{aligned} j_{11} &= F_1(\mathbf{1}) + k_{\text{intra}} F_1(\mathbf{1}) j_{11} + q\frac{k_{\text{inter}}}{m-1}(j_{21} + j_{31} + \cdots + j_{m1}) \\ j_{12} &= q + q\, k_{\text{intra}}\, j_{11} + q\frac{k_{\text{inter}}}{m-1}(j_{21} + j_{31} + \cdots + j_{m1}) \\ &\vdots \end{aligned}$$

$$
\begin{aligned}
j_{1m} &= q + q\, k_{\text{intra}}\, j_{11} + q\frac{k_{\text{inter}}}{m-1}(j_{21} + j_{31} + \cdots + j_{m1}) \\
j_{21} &= q\, k_{\text{intra}}\, j_{22} + q\frac{k_{\text{inter}}}{m-1}(j_{12} + j_{32} + \cdots + j_{m2}) \\
j_{22} &= k_{\text{intra}} F_2(\mathbf{1}) j_{22} + q\frac{k_{\text{inter}}}{m-1}(j_{12} + j_{32} + \cdots + j_{m2}) \\
&\vdots \\
j_{2m} &= q\, k_{\text{intra}}\, j_{22} + q\frac{k_{\text{inter}}}{m-1}(j_{12} + j_{32} + \cdots + j_{m2}) \\
&\vdots \\
j_{m1} &= q\, k_{\text{intra}}\, j_{mm} + q\frac{k_{\text{inter}}}{m-1}(j_{1m} + j_{2m} + \cdots + j_{(m-1)m}) \\
j_{m2} &= q\, k_{\text{intra}}\, j_{mm} + q\frac{k_{\text{inter}}}{m-1}(j_{1m} + j_{2m} + \cdots + j_{(m-1)m}) \\
&\vdots \\
j_{mm} &= k_{\text{intra}} F_m(\mathbf{1}) j_{mm} + q\frac{k_{\text{inter}}}{m-1}(j_{1m} + j_{2m} + \cdots + j_{(m-1)m})
\end{aligned}
\tag{3.27}
$$

Since $F_1(\mathbf{1}) = F_i(\mathbf{1}) = e^{-k_{\text{inter}}}(1-q) + q$ for all i, we can sum all equations for j_{ss} obtaining:

$$
\sum_{s=1}^{m} j_{ss} = F_1(\mathbf{1}) + k_{\text{intra}} F_1(\mathbf{1})\Big(\sum_{s=1}^{m} j_{ss}\Big) + q\frac{k_{\text{inter}}}{m-1}\Big(\sum_{\substack{s,t=1 \\ s\neq t}}^{m} j_{st}\Big) \tag{3.28}
$$

Summing for all equations j_{st} for $s \neq t$ we obtain

$$
\sum_{\substack{s,t=1 \\ s\neq t}}^{m} j_{st} = (m-1)q + (m-1)\, q\, k_{\text{intra}}\Big(\sum_{s=1}^{m} j_{ss}\Big) + q\, k_{\text{inter}}\Big(\sum_{\substack{s,t=1 \\ s\neq t}}^{m} j_{st}\Big)
$$

$$
\Rightarrow \sum_{\substack{s,t=1 \\ s\neq t}}^{m} j_{st} = \frac{(m-1)q + (m-1)\, q\, k_{\text{intra}}\Big(\sum_{s=1}^{m} j_{ss}\Big)}{1 - q\, k_{\text{inter}}} \tag{3.29}
$$

And by substituting 3.29 into 3.28, we obtain

$$
j_{11} + \cdots + j_{mm} = F_1(\mathbf{1}) + k_{\text{intra}} F_1(\mathbf{1}) + \frac{q^2\, k_{\text{inter}}}{1 - q\, k_{\text{inter}}} + \frac{q^2\, k_{\text{intra}}\, k_{\text{inter}}(j_{11} + \cdots + j_{mm})}{1 - q\, k_{\text{inter}}}
$$

$$\Rightarrow j_{11}+\cdots+j_{mm}=\frac{F_1(\mathbf{1})(1-q\,k_{\text{inter}})+q^2\,k_{\text{inter}}}{(1-k_{\text{intra}}F_1(\mathbf{1}))(1-q\,k_{\text{inter}})-q^2\,k_{\text{intra}}\,k_{\text{inter}}} \tag{3.30}$$

leading to the following critical occupation probability of interconnected nodes (in which the average component size diverges):

$$\begin{aligned} q_c &= \frac{-b+\sqrt{b^2-4ac}}{2a} \\ \text{where} \quad a &= k_{\text{intra}}k_{\text{inter}}e^{-k_{\text{inter}}} \\ b &= k_{\text{intra}}+k_{\text{inter}}-k_{\text{intra}}e^{-k_{\text{inter}}}-k_{\text{intra}}k_{\text{inter}}e^{-k_{\text{inter}}} \\ c &= k_{\text{intra}}e^{-k_{\text{inter}}}-1 \end{aligned} \tag{3.31}$$

From these solutions, we obtain the critical occupation probability p_c, using Eq. 3.19.

Finally, once the giant component emerges ($p > p_c$), $J_i(\mathbf{x})$ gives the distribution of i-nodes which are *not* in the giant component, which means that $J_i(\mathbf{1})$ is equal to the fraction of i-nodes which is not occupied by the giant component. The fraction of i-nodes belonging to the giant component, S_i, is therefore given by

$$S_i = 1-J_i(\mathbf{1}) = F_i(\mathbf{1})-F_i(u_{1i},u_{2i},\ldots,u_{mi}) \tag{3.32}$$

where u_{ji} is the probability that an i-node arrived by following a randomly chosen ij-link is not in the giant component. This probability is precisely equal to the probability that none of the neighbours of that node are themselves members of the giant component, and hence it satisfies the self-consistency condition

$$u_{ji} = 1-F_i(\mathbf{1})+F_{ji}(u_{1j},u_{2j},\ldots,u_{mj}) \tag{3.33}$$

Due to symmetry, the size of the giant component, S, equals to the fraction of occupied nodes from module i in the giant component, S_i, and thus the u_{ji} which satisfies equation 3.33 is $u_{ji} = 1-S$ yielding

$$S = e^{-k_{\text{inter}}}(1-q)(1-e^{-k_{\text{intra}}S})+q(1-e^{-(k_{\text{intra}}S+k_{\text{inter}}S)}). \tag{3.34}$$

3.4 Results

We confirm the analytical results obtained above with extensive numerical simulation of ER modular networks of size $N = 600\,000$, which are generated as explained in section 3.2. In Fig. 3.3 we show the percolation threshold as a function of the number of modules m where

the mean degree is kept fixed $k = 4$ and $\alpha = 100$. Green points (blue points) represent results obtained from at least 1000 simulation runs where nodes are randomly removed (where attacking interconnected nodes). Red solid lines represent analytical result obtained in 3.31. Note that apart from this solution, the trivial solution $q_c = 0$ always exists (and represented by a black line), meaning that no interconnected nodes remain in the network and therefore the modules are isolated. In the numerical simulation, at every iteration we remove a fraction 0.01 of the nodes, and estimate the percolation threshold as the point where the second largest component obtains its maximum.

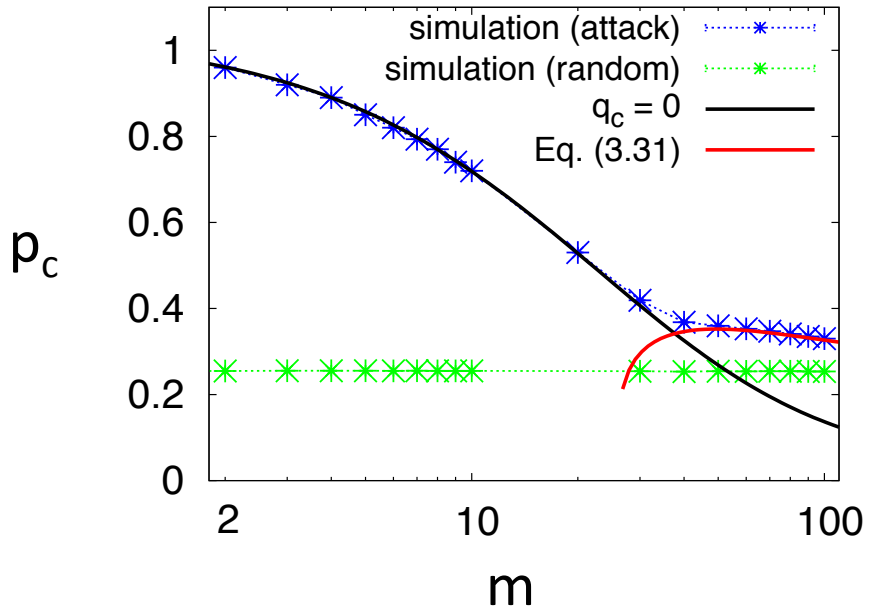


Figure 3.3: Percolation threshold, p_c as a function of the number of modules, m calculated for networks with fixed mean degree $k = 4$ and $\alpha = 100$. Simulation points and error bars (if larger than the marker size) obtained from at least 1000 simulation runs of networks of size $N = 600000$. Black line represents the trivial analytical solution $q_c = 0$ and red line represents the analytical solution obtained in 3.31

Let m^* be the transition point where the two analytical solutions (red and black lines in Fig. 3.3) cross each other. In the regime where $m < m^*$ the only physical solution is $q_c = 0$ meaning that only the removal of all the interconnected nodes breaks down the giant component. The percolation threshold in this case is much higher than for random node removal indicating that the network is extremely fragile due to the inter-connected nodes gluing the system together. As the number of modules and nodes connecting them increases, the network is becoming more robust exhibiting percolation properties similar to single-module networks. For $m > m^*$, it is

enough to remove some of the interconnected nodes to break the system, i.e. the modules' connectivity is affected even before the giant component disappear. This is due to the fact that for $m < m^*$ the attack on interconnected nodes mainly breaks the connectivity between the modules leaving them connected internally, where for $m < m^*$ the interconnected nodes play an important role also in the internal structure of modules and therefore there is no need to remove all of them in order to break down the giant component.

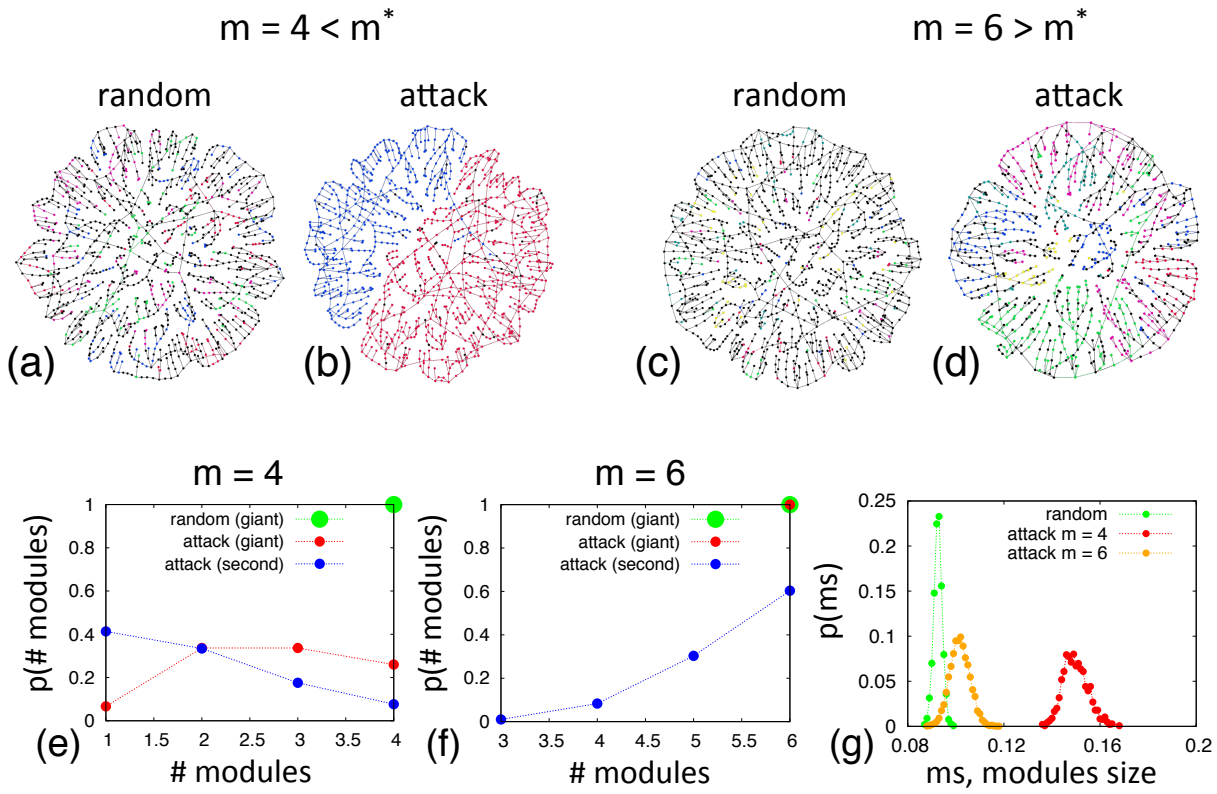


Figure 3.4: Visualisation is shown for networks of size $N = 12000$ with mean degree $k = 4$ and $\alpha = 10$, at the point where the giant component contains 10% of the nodes ($S = 0.1$) (a),(c) for random node removal, (b),(d) for attack on interconnected nodes. (e)-(f) Distribution of the number of modules in the giant component and second largest component at $S = 0.1$. A module is considered to be part of a component if at least one of its nodes are part of the component. (g) Distribution of the size of modules in the giant component at $S = 0.1$, normalized by the initial module size. Note that in (g), the size of modules is measured by reconstructing the graph of each module in the giant component, and counting its number of nodes in this graph. In other words, interconnected nodes that have been detached from their original module are not considered. Results obtained by at least 1000 simulation runs of networks of size $N = 600000$ with mean degree $k = 4$.

In order to illustrate this effect, in Fig. 3.4 we visualise the giant component at $S = 0.1$ (i.e. when the giant component contains 10% of the nodes, which is close to total collapse) with

interconnected nodes shown in black and all other nodes coloured according to the module they belong to. Thus, we run the numerical simulation starting close to $S = 0.1$ with very small steps of 0.001 and when the giant component first goes below 0.1, we count the number of components which appear in it (Fig. 3.4(e)-(f)) and their size (Fig. 3.4(g)). We repeat this computation 1000 times and present the obtained distribution from all runs. In addition, we randomly choose one simulation run, and visualise the obtained network.

For a network with $m = 4 < m^*$, random node failure destroys the internal structure of the modules evenly, see Fig. 3.4(a). In this random failure case, all the modules always appear in the giant component (i.e. there is always at least one node from each module in the giant component) as shown in Fig. 3.4(e), and the size of modules is very narrowly distributed, see Fig. 3.4(g). In contrast to random failure, when attacking the interconnected nodes (at $S = 0.1$), see Fig. 3.4(b), not all the modules remain in the giant component (for example, in Fig. 3.4(b) there are only two of them). However, the modules that do remain, are almost intact, containing 14.6% of their initial nodes, significantly more than in the random case.

For $m > m^*$ the interconnected nodes play an important role in the internal structure of modules, but nevertheless, the attack still leaves them more complete than in the case of random removal, see Fig. 3.4(c)-(d). Furthermore, in the case of attack, usually *all* modules appear in the giant component (see Fig. 3.4(f)), and thus their relative size is smaller compared to the $m < m^*$ case (see Fig. 3.4(g)). As m increases, the difference between attack and random case becomes smaller, and as a result the percolation threshold converges to the one obtained for random failure.

In Fig. 3.5 we show the fraction of nodes in the largest cluster, S, and second largest cluster S_{second} as the fraction of occupied nodes p increases. In the regime $m < m^*$, the attack on interconnected nodes has a weak effect on the internal structure of the modules, and their removal results in an abrupt decrease in the size of the giant component. In addition, while for $m = 100 > m^*$ (Fig. 3.5(b)) we observe a regular second order percolation transition characterized by the continuous decrease of S and the sharp peak in S_{second}, the case of $m < m^*$ demonstrates an abrupt, first order transitions. The reason is that the second largest cluster contains large connected subgraphs corresponding to modules who "dropped" from the giant component, see Fig. 3.4(e). Therefore, with the emergence of the giant component, these modules become part of it, leading to a sudden drop in the size of the second largest cluster.

In Fig. 3.6(a), we show the critical number of modules, m^*, as a function of α for networks with mean degree $k = 4$. It is seen that m^* is increasing with α, and the percolation threshold at this point is ${p_c}^* \approx 0.3417$ independent of α, meaning the transition takes place at a fixed

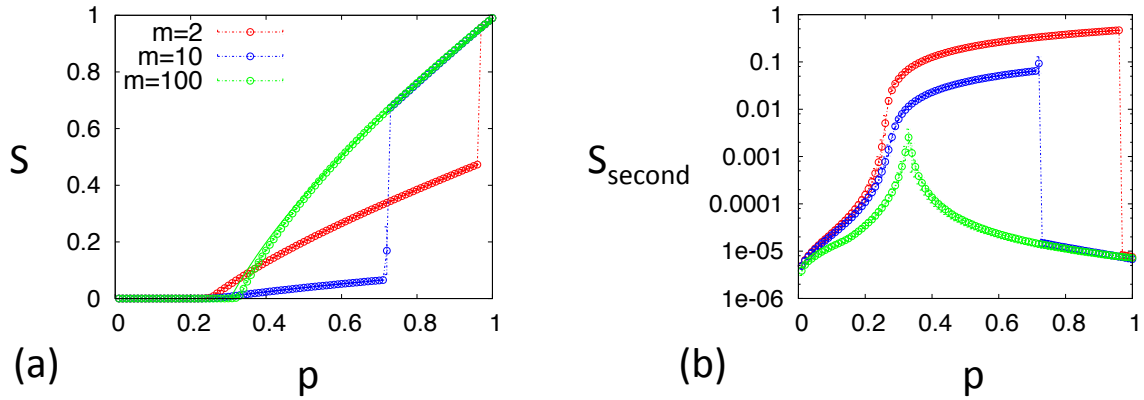


Figure 3.5: Fraction of nodes in the largest cluster, S and fraction of nodes in the second largest cluster, S_{second}, as a function of occupation probability, p, for networks with mean degree $k = 4$ and $\alpha = 100$. Simulation points and error bars (if larger than the marker size) obtained from at least 1000 simulation runs of networks of size $N = 600\,000$. Solid lines represent the analytical result obtained in 3.34 and computed using rootSolve [214], an R package [215].

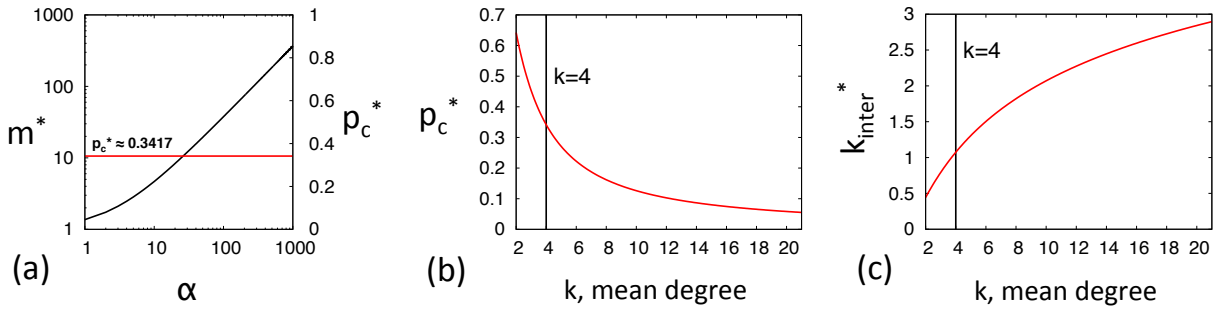


Figure 3.6: (a) Critical number of modules m^*, defined as the point where the analytical solutions (black and red lines in Fig. 3.3) cross each other, as a function of α in networks with mean degree $k = 4$. ${p_c}^*$ and ${k_{\text{inter}}}^*$ are the percolation threshold and the mean inter-degree at this point respectively. (b)-(c) Change in the critical percolation threshold, p_c^*, and the critical concentration of inter-nodes, ${k_{\text{inter}}}^*$ as a function of the mean degree k. Both quantities are independent of α.

inter-module average degree ${k_{\text{inter}}}^* = -\ln({p_c}^*) \approx 1.0738$. In Fig. 3.6(b)-(c), we show how the critical percolation threshold ${p_c}^*$ and the critical mean inter-degree ${k_{\text{inter}}}^*$ are changing with the mean-degree k. Note that in Fig. 3.6(a) the results are for networks with fixed mean degree $k = 4$, which is marked with a black vertical line in Fig. 3.6(b)-(c).

In order to further demonstrate the transition in the p_c behaviour, in Fig. 3.7 we show the percolation threshold as a function of k_{inter} for networks with mean intra-degree $k_{\text{intra}} = 2$ and

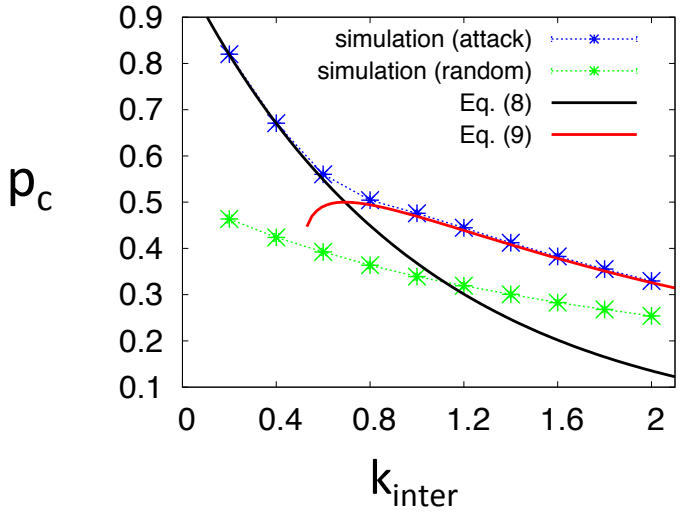


Figure 3.7: Percolation threshold, p_c as a function of the mean inter-degree, k_{inter} calculated for networks with $m = 10$ modules, mean intra-degree $k_{\text{intra}} = 2$ and $\alpha = 100$. Simulation points and error bars (if larger than the marker size) obtained from at least 1000 simulation runs of networks of size $N = 600\,000$. The black line represents the trivial analytical solution $q_c = 0$ and the red line represents the analytical solution obtained in 3.31.

number of modules $m = 10$ fixed. Here we see a similar transition in p_c as before, but the critical point is now a function of the concentration of interconnected nodes. At a critical $k^*_{\text{inter}} = k_{\text{inter}} \approx 0.693$, p_c changes from $q_c = 0$ behaviour to Eq. 3.31 behaviour.

Finally, in modular structures the interconnected nodes have high betweenness centrality, defined as the fraction of shortest paths in the network that pass through a node [97, 98], and thus, our framework also provides an analytical tool of studying attacks on high betweenness centrality nodes, where only numerical simulations currently exist that suggest such an attack is one of the most harmful attack strategies [125, 201]. Figure 3.8 compares the betweenness centrality of nodes with inter-module connections (called inter-nodes) and nodes with only intra-module connections (called intra-nodes) for networks of size $N = 100\,000$ with $m = 10$ modules. First, we show that the average betweenness centrality of interconnected nodes is significantly higher than for nodes without interconnections in networks with mean intra-degree $k_{\text{intra}} = 2$ and a varying number of interconnections, see Fig. 3.8(a). Then, for $k_{\text{inter}} = 2$, we show that the betweenness centrality distribution of interconnected nodes has a broader tail, meaning that interconnected nodes are much more likely to have high betweenness centrality. Thus, our analytical results of attack on interconnected nodes can be regarded as a theory for attacking high betweenness nodes.

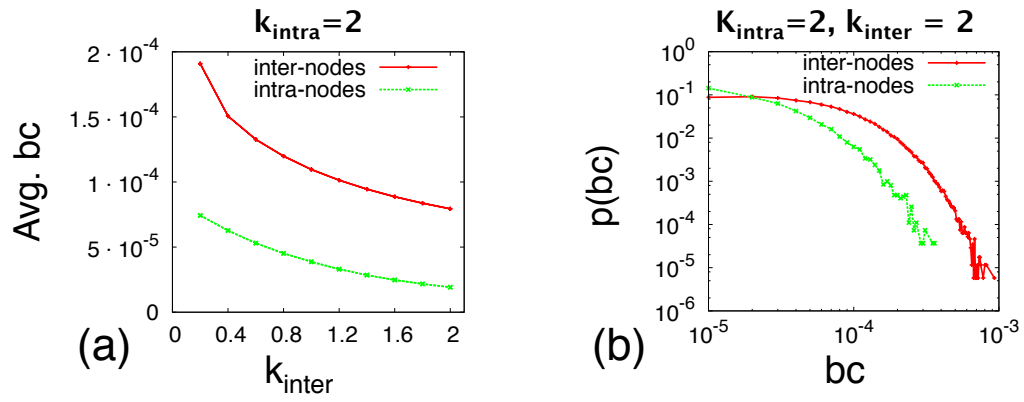


Figure 3.8: Betweenness centrality of inter-nodes (nodes that have at least one interconnection) and intra-nodes (nodes with only intraconnections) in networks of size $N = 100000$ with $m = 10$ modules. (a) Mean betweenness centrality as a function of k_{inter} in networks with $k_{\text{intra}} = 2$. (b) Distribution of betweenness centrality in networks with $k_{\text{intra}} = k_{\text{inter}} = 2$.

Although throughout this chapter we have only considered modular ER networks, in the following we would like to discuss numerical results for scale-free networks. In Fig. 3.9(a)-(b) we show numerical results for scale-free networks with scaling exponent 2.5 and 3.5 respectively, generated using the configuration model. In particular, similarly to the construction of the modular ER networks described in section 3.2, we start by building m scale-free networks of size $\frac{N}{m}$ using the R method "degree.sequence.game" which implements an algorithm to generate undirected, connected simple graphs (with no self-edges and multi-edges) with a given degree sequence [8, 158, 221]. We create the degree sequence (for each module) by drawing $\frac{N}{m}$ random numbers from a power-law degree distribution with a given scaling exponent [66], and then normalise each number such that their sum would equal to twice the number of required edges, $\frac{M_{\text{intra}}}{m}$. Once we have built each module internally, we connect between the modules by creating another degree sequence as before, and randomly connecting M_{inter} couples of nodes from two different modules, with sampling probability of a node proportional to its inter-degree obtained from the degree sequence.

For scale-free networks with scaling exponent 2.5, see Fig. 3.9(a), random node failure results in a very small percolation threshold, which we suspect (and also observe in other numerical results not shown here) that is converging to zero as the size of modules $\frac{N}{m}$ is increasing, in agreement with the analytical result obtained in section 3.3.2 where modular networks behaves similarly to single-module networks in the presence of random failure. The behaviour is much different, however, when considering attack on interconnected nodes (blue points). In this case, for scaling exponent 3.5, see Fig. 3.9(b), we observe a similar qualitative behaviour as in Fig. 3.3, where at

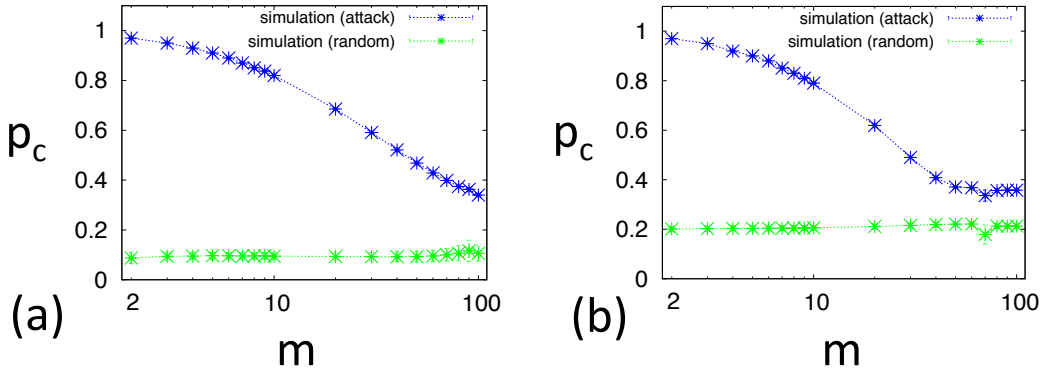


Figure 3.9: Percolation threshold, p_c as a function of the number of modules, m calculated for modular scale-free networks with scaling exponent 2.5 (a) and 3.5 (b), mean degree $k = 4$ and $\alpha = 100$. Simulation points and error bars (if larger than the marker size) obtained from at least 1000 simulation runs of networks of size $N = 600\,000$.

some critical m, the percolation threshold is suddenly changing its behaviour. We hypothesise that this critical m, just like m^* is marking the point where the modules are affected internally as a result of the attack, and thus the system collapses before removing all the interconnected nodes. This behaviour is not observed for scaling exponent 2.5 (Fig. 3.9(a)) since in this case the critical m would be very high (outside the range shown in the image). This is because modules with scale-free structure with scaling exponent 2.5 are very robust, and thus in order to damage their internal structure, there needed to be a very high concentration of interconnected nodes, which happens only at larger m.

3.5 Conclusion

To summarise, in this chapter we presented a theoretical framework to study the resilience of modular networks to attack on interconnected nodes, those connecting between the modules. The robustness of heterogeneous systems consisting of interconnected modules (or networks) has received much attention in recent years, due to their importance for the functionality of biological systems, including brain networks. However, only recently, with the new analytical tools developed in the exploding body of work on coupled networks, are we able to systematically study modularity structure, and its effect on the vulnerability of networks.

We discover two percolation regimes separated by a critical concentration of interconnections between modules, above which the internal structure of each module is inseparable from the system as a whole. In particular, when the number of modules and interconnections between them

is below a certain threshold, the only way to break the system is removing all the interconnections since the modules themselves are unaffected by this kind of attack. However, above a certain threshold, the removal of interconnected nodes already breaks the modules internally which helps to break the whole system, i.e. it is enough to remove some interconnected nodes such that the whole system collapses.

Since in both percolation regimes the percolation threshold obtained from attacking the interconnected nodes is higher than the case of random failure, our study offers an efficient immunisation approach, where epidemic spreading can be prevented at a low cost by immunising interconnected nodes. At the regime where the interconnected nodes are not playing a significant roles in the internal structure of modules, this can be done at an extremely low cost as the percolation threshold is very high. Thus, in geographically distant social networks, it is worth vaccinating people that link between different communities such as businessmen traveling a lot between countries (which are likely to found at airports for example).

Our approach can also be used to study analytically attacks on high betweenness centrality nodes, which in modular structures, correspond to interconnected nodes. Such attacks, which have only been studied numerically so far, are considered to be among the most harmful attack strategies.

Finally, our results provide essential insights for the interplay between the number of modules (or sub-networks), m, their intra and inter-connectivity patterns α, and the percolation threshold, which could be utilised to find an optimal partition of large-scale systems into sub-networks in order to maintain both robustness and scalability.

There are many interesting directions in which to extend the analytical framework developed here, and examine the stability of the two percolation regimes that we found. First, following the numerical results we presented in the last figure, scale-free networks seem to exhibit the same qualitative behaviour as homogeneous ER networks, a result which we would like to validate analytically. Second, the case of more heterogeneous networks where each module can be of different size, degree distribution and etc, is very important, but would require a more refine analysis than the one considered here. In particular, symmetry arguments as the one used to derive Eq. 3.34 will need to be revised. Finally, with the new analytical tools presented here, it would be interesting to consider other types of attacks, mainly those exist in biological networks where modular structure is so crucial to the global collective behaviour of the network.

CHAPTER FOUR

EPIDEMIC SPREADING ON COUPLED ADAPTIVE NETWORKS

Most real-world networks are adaptive, they have the ability to adapt their topology dynamically in response to the dynamic states of nodes. However, networks rarely live in isolation, but instead interact with or depend on other networks, resulting in coupled adaptive networks. In order to understand the implications of interconnection between otherwise independently adaptive networks, in this chapter, we present an analytical framework to study epidemic spreading in coupled adaptive networks. The chapter is based on the author's work [204].

4.1 Introduction

Adaptive networks, which combine topological evolution of the network with dynamics on the network, are ubiquitous across disciplines [115]. Social, biological and technological networks all have the ability to adapt their topology dynamically in response to the dynamic states of nodes. For example, in distribution networks, such as the vascular system, transportation networks, power grids and the Internet, the nodes and links are subject to high loads which can lead to failures e.g. traffic jam, electrical line failure. But repeating failures can in turn lead to the formation of new nodes and links to relieve this load. In the vascular system, for instance, increasing shear stress on the arteries as a result of an increasing flow velocity, can lead to the formation of new arteries in a process called arteriogenesis, in order to prevent a dangerous restriction in blood supply (ischemia) [200]. The mutual interaction between an evolving network

topology and the nodes' dynamics has been shown to give rise to rich behaviours including self-organising criticality [48, 154], spontaneous division of labor (the emergence of distinct classes of nodes from an initially homogeneous population) [83, 133] and cooperation [87, 225, 251].

However, most real-world networks are not isolated but interact with or depend on other networks, resulting in *coupled adaptive networks* in which there are links between otherwise independent networks. For example, brain networks consisting of synaptic connections linking neural units can be divided into separate cortical areas coupled together through interregional pathways [219]. Technical distribution networks are often divided into subsystems where each subsystem is adaptive and interconnected to other subsystems (e.g. data centres). Finally, distant populations are linked via people traveling (e.g. by flights) between the networks, creating interacting networks [146]. It is therefore essential to understand how interlinks between adaptive networks are affecting their behaviour. We used the term "coupled" in this chapter, and not modular as in previous chapter, as we consider the case of two adaptive networks where the interlinks connecting them are behaving differently from the links within each network. But apart from this, they are mathematically the same, both representing a system where multiple networks are interacting with each other.

In this chapter, we present an analytical framework, based on nonlinear ordinary differential equations, to study coupled adaptive networks, using which we demonstrate that inter-network links between adaptive networks can give rise to nontrivial behaviours, and thus must be taken in consideration. By applying bifurcation analysis, we study the stable equilibria of an epidemic spreading taking place on a system of two adaptive networks coupled together by inter-network links, which can also spread the disease. We find that stability is increased by increasing the number of inter-network links (i.e. higher coupling), in the sense that the range of parameters over which both endemic and healthy states coexist (both stable states are reachable depending on the initial conditions) becomes smaller. And perhaps more importantly, we find a new stable state that does not appear in the case of a single adaptive network but only in the case of weakly coupled networks. While in single networks, the epidemic either breaks to the entire network or dies out, interconnections can result in a localise epidemic breakout only in the coupled nodes that does not break out into the wider network.

The rest of this chapter is organised as follows: in section 4.2 we present our model of SIS epidemic spreading in coupled adaptive networks. In section 4.3 we present an analytical formalism, based on nonlinear differential equations, to study the time evolution and stationary disease prevalence in each network. Then, in section 4.4, we provide details on the computer simulation used to verify the analytical predictions. In section 4.5 we compare the analytical and numerical results, and discuss the limitations of our analytical framework, and finally, in

section 4.6 we summarise our findings and discuss their implications.

4.2 Model

We consider two networks A and B with the same number of nodes $N_A = N_B = N$, where a randomly-chosen fraction of nodes p_{coup} (the *coupling probability*) from each network are coupled together through randomly-assigned inter-network connections. Each coupled node has exactly one interconnection to a randomly chosen coupled node in the other network. This is a similar model to the one of partially interdependent networks studied by Parshani *et al.* [180], but in our model the inter-network links are regular connectivity links instead of dependency links, representing interaction between networks, such that an epidemic or rumor that is spreading in one network, for example, can spread to the other. Though simple, this model already gives rise to nontrivial behaviours, and can be extended to account for more general cases, for example by using an arbitrary inter-network degree distribution [82, 198].

The dynamics we consider is an extension of the well established adaptive SIS epidemic spreading model of Gross *et al.* [112]. In this model, similarly to SIS epidemic spreading in static (nonadaptive) networks (see section 2.3.2), nodes can be in either one of two states – susceptible or infected – where susceptible nodes can get infected from their infected neighbors and infected nodes can recover and become susceptible again. In addition, the network is adaptive in the sense that susceptible nodes can rewire their links from infected neighbors to randomly selected susceptible ones. In a human-contact network, for example, this corresponds to not meeting a friend one knows to be ill. This dynamics has been shown [112] to lead to the emergence of new epidemic thresholds and the coexistence of two stable equilibria where both endemic and disease-free states are attractive while in static networks only one stable equilibrium is observed. Recent studies consider more realistic scenarios extending Gross *et al.*'s model such as susceptible-infected-recovered-susceptible (SIRS) epidemic dynamics where infected nodes are recovered (and temporarily immune to the disease) before they become susceptible again [207], susceptible-infected-vaccinated (SIV) dynamics where vaccination occurs in Poisson-distributed pulses [208] and the effect of different rewiring rules [243]. These models result in similar dynamics that is not observed in static network model.

In our model, each of the networks, A and B, follows the dynamics rules of Gross *et al.*'s model, and in addition, the coupled nodes can also get infected from their inter-network neighbours. However, the inter-network links are permanent (nonrewireable), accounting for a common situation in coupled networks: while the individual networks are independently adaptive, their dependencies – the links between them – are often permanent (nonadaptive). This limits an

individual network's ability to adapt in the face of challenges, since it can change its own topology but not its dependence on other networks. In geographically distant social networks, for example, long-distance links are often family links which are not subject to rewiring compared to short-distance friend and acquaintance links. The inter-network links between the networks create an interesting situation where nodes in each network change their links according to the states of their neighbors in the same network, but are also affected by the states of the nodes in the coupled network. Using the human-contact example again, there are some people one cannot avoid even if one knows them to be ill.

The dynamics rules of the whole system can be described as follows. Starting with two networks of susceptible nodes, we introduce a seed proportion of infected nodes in each network $0 \leq \varepsilon_A, \varepsilon_B \leq 1$. Then at every time step infected nodes can pass the disease to their intra and inter-network neighbors at some rate (the *infection rate*), while they recover and become susceptible again at some other rate α (the *recovery rate*). Let β_A, β_B, β_{inter} be the infection rates within network A, B and between the networks respectively. For every intra-network link connecting an infected with a susceptible in network A (respectively network B), the susceptible node becomes infected with infection rate β_A (β_B). Also, for every inter-network link connecting an infected node from network A (network B) with a susceptible from network network B (network A), the susceptible node becomes infected with infection rate β_{inter}. In addition, we allow susceptible nodes to protect themselves by rewiring their intra-network links. For every intra-network link connecting a susceptible node with an infected node in network A (network B), the susceptible node rewires the link with rewiring probability γ: it breaks the link to the infected node and forms a new intra-network link to another randomly-selected susceptible node, where double- and self-connections are prohibited, thus conserving the number of intra and inter-network links.

4.3 Formalism

We proceeded by studying the time evolution and stationary disease prevalence arising from the local transformation rules described above. Let $[I_A]$ and $[S_A]$ denote the fraction of infected and susceptible nodes in network A respectively, and similarly $[I_B]$ and $[S_B]$ in network B. By the conservation of the number of nodes we obtain

$$[I_A]+[S_A]=1$$
$$[I_B]+[S_B]=1$$

Let $[S_A S_A], [I_A I_A], [I_A S_A]$ denote the densities per node of intra-network links in network

A connecting susceptible to susceptible, infected to infected and infected with susceptible respectively. Note that $[I_A S_A]$ accounts for all the links connecting either susceptible to infected or infected to susceptible in network A. In other words, we do not distinguish between $[I_A S_A]$ and $[S_A I_A]$, or between $[I_B S_B]$ and $[S_B I_B]$. By the conservation of the number of intra-network links we obtain

$$[S_A S_A]+[I_A I_A]+[I_A S_A]=\frac{\langle k_A\rangle}{2}$$
$$[S_B S_B]+[I_B I_B]+[I_B S_B]=\frac{\langle k_B\rangle}{2}$$

where $\langle k_A\rangle$, $\langle k_B\rangle$ are the mean degree of nodes in networks A and B respectively. Note that the mean degrees only account for intra-network links while ignoring inter-network links. The latter are considered in the coupling probability p_{coup}.

Finally, let $[S_A S_B], [I_A I_B], [I_A S_B], [I_B S_A]$ denote the densities per node of inter-network links connecting susceptible A node to susceptible B node, infected A node to infected B node, infected A node to susceptible B node and infected B node to susceptible A node respectively. By the conservation of the number of inter-network links we obtain

$$[S_A S_B]+[I_A I_B]+[I_A S_B]+[I_B S_A]=p_{\text{coup}}$$

where p_{coup}, the coupling probability, is the fraction of nodes that are connected to nodes in the other network through inter-network links.

We may now derive a system of ordinary differential equations (ODEs) describing the time evolution of the model. The time evolution of the fraction of infected nodes in each network is given by

$$\frac{d[I_A]}{dt}=\beta_A[I_A S_A]+\beta_{\text{inter}}[I_B S_A]-\alpha[I_A] \tag{4.1}$$

$$\frac{d[I_B]}{dt}=\beta_B[I_B S_B]+\beta_{\text{inter}}[I_A S_B]-\alpha[I_B] \tag{4.2}$$

The first and second terms in Eq. 4.1 describe the infection of susceptible nodes in network A due to intra- and inter-network links respectively, while the third term describes recovery, and similarly for network B.

Since the equations describing the time evolution of nodes depending on the time evolution of links, we proceed by developing the latter. The time evolution of the densities of intra-network

links connecting infected to infected per node is given by

$$\frac{d[I_A I_A]}{dt} = \beta_A[I_A S_A] + 2\beta_A[I_A S_A I_A] + \beta_{inter}[I_A S_A I_B] - 2\alpha[I_A I_A] \tag{4.3}$$

$$\frac{d[I_B I_B]}{dt} = \beta_B[I_B S_B] + 2\beta_B[I_B S_B I_B] + \beta_{inter}[I_B S_B I_A] - 2\alpha[I_B I_B] \tag{4.4}$$

The first and second terms in Eq. 4.3 correspond to the conversion of $S_A I_A$ links into $I_A I_A$ links as a result of new infections through intra-network connections, while the third describes the conversion of these links as a result of infections through inter-network links. The fourth term represents the conversion of $I_A I_A$ links into $S_A I_A$ links as a result of recovery. Similarly, we can write the evolution of links connecting susceptible to susceptible as

$$\frac{d[S_A S_A]}{dt} = (\alpha+\gamma)[I_A S_A] - \beta_A[S_A S_A I_A] - \beta_{inter}[S_A S_A I_B] \tag{4.5}$$

$$\frac{d[S_B S_B]}{dt} = (\alpha+\gamma)[I_B S_B] - \beta_B[S_B S_B I_B] - \beta_{inter}[S_B S_B I_A] \tag{4.6}$$

The first and second terms in Eq. 4.5 correspond to the conversion of $I_A S_A$ links into $S_A S_A$ links as a result of recovery and rewiring, and the third and fourth terms describe the conversion of $S_A S_A$ links into $I_A S_A$ links as a result of infection through intra- and inter-networks links respectively.

Finally, we describe the time evolution of the densities of inter-network links per node

$$\frac{d[I_A I_B]}{dt} = \beta_{inter}[I_A S_B] + \beta_{inter}[I_B S_A] + \beta_A[I_A S_A I_B] + \beta_B[I_B S_B I_A] - 2\alpha[I_A I_B] \tag{4.7}$$

$$\frac{d[S_A S_B]}{dt} = \alpha[I_A S_B] + \alpha[I_B S_A] - \beta_A[I_A S_A S_B] - \beta_B[I_B S_B S_A] \tag{4.8}$$

$$\frac{d[I_A S_B]}{dt} = \alpha[I_A I_B] + \beta_A[I_A S_A S_B] - \alpha[I_A S_B] - \beta_{inter}[I_A S_B] - \beta_B[I_B S_B I_A] \tag{4.9}$$

The first and second terms in Eq. 4.7 correspond to the conversion of inter-links connecting susceptible to infected into $I_A I_B$ links as a result of infection through inter-network connections, while the third and fourth describe the same conversion due to infection through intra-network connection. The last term corresponds to the recovery of either an A or B node. Similarly, the first and second terms in Eq. 4.8 correspond to recovery while last two terms correspond

to infection through intra-network links. Note that infection due to inter-network links is not possible here since both nodes are susceptible. Finally, the first and second terms in Eq. 4.9 describe the gain of inter-network links connecting susceptible A node with infected B node due to recovery of a B node and the infection of an A node through intra-network neighbour. The last three terms describe the loss of $I_A S_B$ links due to recovery of an A node and the infection of a B node through inter- and intra-network links respectively.

Similarly to equations 4.1-4.2, where the time evolution of nodes depend on the time evolution of links, here we obtain that the equations describing the time evolution of links densities depend on the time evolution of triplets. In other words, in order to know how the density of *IS*-links (connecting an infected node to a susceptible node) for example is changing in time, we need to know how the density of triplets involving *IS*-links, i.e. *IIS*, *SIS*, *ISS* and *ISI* triplets, is changing in time. This is a typical problem in such contact processes, describing the transmission of distinct properties of nodes via the links of a network [116]: the dynamics of subgraphs with n nodes depends on the dynamics of the larger subgraphs in which they are contained, resulting in an infinite cascade of ODEs. A common approach to solve the infinite cascade of ODEs is to simply truncate it at some order n, meaning that instead of writing the dynamical equations for subgraphs with n nodes, they are approximated by the density the of the smaller subgraphs constituting them [116].

In the following we are using the pair approximation method [112, 114, 137], where the density of triplets is expressed in terms of links and nodes. For example, given an *IS*-link in network A (a link connecting a susceptible node to an infected node in network A), which is of density $[I_A S_A]$, we can approximate $[I_A S_A I_A]$ by deriving the expected number of infected nodes which are connected to a susceptible node at the end of an *IS*-link. Such susceptible node has by definition $\langle q_A \rangle$ additional links, where $\langle q_A \rangle$ is the mean excess degree of network A (see section 2.3.1). Each of these links is an *SI*-link with probability $\frac{[I_A S_A]}{\langle k_A \rangle [S_A]}$. Therefore, the density of $I_A S_A I_A$ triplets can be approximated by $\frac{\langle q_A \rangle [I_A S_A][I_A S_A]}{\langle k_A \rangle [S_A]}$. For the sake of simplicity, we will only consider Erdős-Rényi (ER) networks, for which the mean excess degree is equal to the mean degree, meaning that $\langle k_A \rangle$ and $\langle q_A \rangle$ are cancelling each other [114, 175]. Extending to other network models is possible by deriving an expression for the mean excess degree (see section 2.3.1).

Similarly, the expected number of infected coupled nodes from network B connecting to an *SI*-link in network A to create a $I_A S_A I_B$ triplet is p_{coup}, which is also the average inter-connectivity, and thus the density of $I_A S_A I_B$ triplets can be approximated by $\frac{[I_A S_A][I_B S_A]}{[S_A]}$. Similarly to how we derived the expression for $[I_A S_A I_A]$ and $[I_A S_A I_B]$, we can now approximate the density of all triplets by

$$[I_A S_A I_A] = \frac{[I_A S_A][I_A S_A]}{2[S_A]} \quad ; \quad [I_B S_B I_B] = \frac{[I_B S_B][I_B S_B]}{2[S_B]}$$
$$[S_A S_A I_A] = \frac{2[S_A S_A][I_A S_A]}{[S_A]} \quad ; \quad [S_B S_B I_B] = \frac{2[S_B S_B][I_B S_B]}{[S_B]}$$
$$[I_A S_A I_B] = \frac{[I_A S_A][I_B S_A]}{[S_A]} \quad ; \quad [I_B S_B I_A] = \frac{[I_B S_B][I_A S_B]}{[S_B]}$$
$$[S_A S_A I_B] = \frac{2[S_A S_A][I_B S_A]}{[S_A]} \quad ; \quad [S_B S_B I_A] = \frac{2[S_B S_B][I_A S_B]}{[S_B]}$$
$$[I_A S_A S_B] = \frac{[I_A S_A][S_A S_B]}{[S_A]} \quad ; \quad [I_B S_B S_A] = \frac{[I_B S_B][S_B S_A]}{[S_B]}$$

This first-order approximation assumes – similarly to the mean-field approximation, where S and I nodes are assumed to be homogeneously distributed in the network – the homogeneous distribution of links of any kind in the network. This assumption might not be valid at the beginning of the spreading process where only a seed of infected nodes exist in the network (thus, *SI*-links for example, are concentrated around this seed), but usually gives a good approximation for the situation at equilibrium. However, in coupled adaptive networks, we find a new equilibrium where the epidemic localise only at the coupled nodes, thus breaking the homogeneous link distribution assumption, resulting in a disagreement between the analytical solution and the computer simulation. In this case, as we show in the results section, we substitute the pair approximation with the densities of triplets extracted straight from the simulation. This indeed solves the problem (thus also proving that the disagreement is rooted at the pair approximation), but is computationally very expensive, which is a major motivation of having the analytical solution in first place. Gross *et al.* [113] suggested a more efficient algorithm to compute the densities of triplets numerically on-demand from short bursts of appropriately initialized numerical simulations.

In addition, there are several analytical approaches that attempt to provide a more accurate approximation than the pair approximation [80]. One of them is to simply truncate the analytical model at higher order, i.e. developing the dynamical equations describing the time evolution of densities of larger subgraphs [151, 187]. However, this is usually very difficult, as the number of equations in the model grows combinatorially with the order of the expansion and the number of states in the model [114]. Another more recent approach, inspired by the computation of epidemic thresholds, assumes a low density of active links (connecting nodes in different states) close to transition points, which allows more careful handling of long range correlations [43]. We chose the pair approximation method for its simplicity, in order to show the potential of our

analytical formalism. More complicated approaches can, however, be applied easily within the framework of our model.

Equations 4.1-4.9 together with the conservation rules and the pair approximation can now be solved given the density of nodes and links at $t = 0$. As stated previously, we start the epidemic spreading by initially infecting a fraction ε_A of nodes in network A and ε_B in network B, and thus, the initial conditions for Eqs. 4.1-4.9 are given by

$$
\begin{aligned}
&[I_A](0)=\varepsilon_A && ; \quad [I_B](0)=\varepsilon_B \\
&[I_A I_A](0)={\varepsilon_A}^2 \frac{\langle k_A \rangle}{2} && ; \quad [I_B I_B](0)={\varepsilon_B}^2 \frac{\langle k_B \rangle}{2} \\
&[S_A S_A](0)=(1-\varepsilon_A)^2 \frac{\langle k_A \rangle}{2} && ; \quad [S_B S_B](0)=(1-\varepsilon_B)^2 \frac{\langle k_B \rangle}{2} \\
&[I_A I_B](0)=p_{\text{coup}}\ \varepsilon_A \varepsilon_B \\
&[S_A S_B](0)=p_{\text{coup}}(1-\varepsilon_A)(1-\varepsilon_B) \\
&[I_A S_B](0)=p_{\text{coup}}\ \varepsilon_A(1-\varepsilon_B)
\end{aligned}
$$

We are interested to study how the stationary disease prevalence, i.e. the stationary points of the analytical model given above, depend on the spreading parameters: infection rate, rewiring rate, and number of internetwork links. In particular, let $x = ([I_A], [I_B], [I_A I_A], [I_B I_B], [S_A S_A], [S_B S_B], [I_A I_B], [S_A S_B], [I_A S_B])$ represent a multidimensional dynamical variable, whose time evolution depends on the parameters $\mu = (\alpha, \beta_A, \beta_B, \beta_{\text{inter}}, \gamma, p_{\text{coup}})$ in the following way

$$\dot{x} = \frac{\partial x}{\partial t} = f(x, \mu) \tag{4.10}$$

When the model is analytically solvable, it is possible to find a closed formula for the stationary points x_0, in which $f(x_0, \mu) = 0$, as a function of μ (i.e. curve of solutions $x(\mu)$ in which $f(x(\mu), \mu) = 0$), and then analyse their stability by exploring the Jacobian at x_0, $J_f(x_0, \mu) = \left(\frac{\partial f_i}{\partial x_j}\right)|_{x=x_0}$. Generally, small perturbations of μ result in slight changes in the phase space, which is spanned by all possible system trajectories $x(t)$. There are, however, certain parameter values at which a qualitative change occurs, i.e. a change in the topology of the phase space, such as the birth or death of a stationary point or a change in stability. Such points in parameter space are called *bifurcation points* at which the system undergoes a qualitative change in its long-term behaviour [142].

However, generally in nonlinear systems, and also in our case, it is hard to obtain closed

formulas for stationary points $x(\mu)$ (the same dynamics is not analytically solvable even in a single network [112]), and therefore in the following we use the dynamical systems analysis software XPPAUT [92], which implements numerical methods for studying nonlinear differential questions. There are many sophisticated numerical methods for exploring the stability of stationary points as the system parameters vary, and their choice depend on the specific problem in hand, but the general idea is the following. First, we need to numerically solve the equation $f(x,\mu_0)=0$ for some μ_0, in order to find a stationary point. If the system has a stable stationary point, then it can be found by numerical integration of $f(x,\mu_0)$, which will converge to the stable point as $t\to\infty$. Otherwise, if the stationary point is known approximately, then one could use Newton's iteration to correct the approximation. Once a stationary point x_0 is found, we are now interested to reveal the whole solution branch as a function of μ. This can be done using parameter continuation – a numerical method for finding the equilibrium curve $f(x,\mu)=0$ that goes through (x_0,μ_0). By the Implicit Function Theorem, if the Jabobian matrix has full rank at (x_0,μ_0), then locally there is a smooth curve M passing through (x_0,μ_0) [142]. We can now approximate M with a desired accuracy using a sequence of points $(x_1,\mu_1),(x_2,\mu_2)\ldots$, which can be found in the following way. Starting with the initial point (x_0,μ_0), the tangent vector to the curve M at this point, v_0, is computed and is used to find a new point $(\tilde{x_1},\tilde{\mu_1})=(x_0,\mu_0)+h_0v_0$, where h_0 is a small step size. The new point $(\tilde{x_1},\tilde{\mu_1})$ is not on the curve (i.e. it is not a solution for $f(x,\mu)=0$), but it is close to it, and can thus be used as an initial guess for the Newton's iteration method, which will eventually converge to a point (x_1,μ_1) on the curve M. This process is repeated, and given that the steps h_i are small enough, M can be found with a high accuracy.

4.4 Simulation

We confirm results computed analytically by explicitly simulating the dynamics over a system of two coupled adaptive networks using the Largenet2 C++ library [252], which provides suitable data structures that allow efficient implementation of asynchronous network simulation. More specifically, the local transformation rules described in section 4.2, can be applied in more than one way. Synchronous (or parallel) updating schemes are the most widely used in simulation of epidemic spreading [183, 184, 185, 186], where the state of all nodes is updated at discrete time steps, i.e. the state of each node at time t depends on the state of the network at time $t-1$. However, in adaptive networks, simultaneous modifications of local network topologies at more than one place may cause conflicting results that are inconsistent with each other, and thus asynchronous continuous updating is the more common choice [199]. Although different updating schemes might result in different time evolutions of the disease, they both generally lead to the same long-term behaviour [7, 100, 106], and moreover, it has been shown that

asynchronous automata networks can emulate any synchronous automata networks, and thus our result here cover the whole class of dynamics that can be observed under the synchronous updating scheme [168, 199].

We implement the widely-used Gillespie's algorithm [105], which was originally invented for the stochastic simulation of chemical reaction networks, and can be described by the four following steps

(i) Given the current state of the network and the events' probabilities (i.e. inter-infection, intra-infection, rewiring and recovery), we determine the next event to be applied. For example, the probability to apply inter-infection is proportional to β_{inter} multiplied by the number of inter-links connecting susceptible node from one network with an infected node from the other network.

(ii) Given the next event to be applied, we randomly choose an appropriate subgraph to which the event can be applied. In the example above, we randomly choose a inter-link connecting susceptible with infected. In Largenet2, this can be done in a constant time unlike other libraries used so far, like networkX, where the complexity would be an order of the number of edges thus significantly slowing down the simulation.

(iii) Apply the event to the subgraph. In the example above, we change the state of the susceptible node to infected.

(iv) Increment the time by a random number drawn from an exponential distribution with mean $\frac{1}{R_{\text{tot}}}$, where R_{tot} is the sum of all events' probabilities. In other words, the algorithm assumes an exponential decay of the probability that an event has not occurred yet.

Note that in contrast to the synchronous step-based simulation, Gillespie's algorithm is event-based meaning that it jumps directly from event to event, skipping over time steps in which no event takes place. Our simulation results were obtained by averaging at least 25 000 values corresponding to the prevalence at equilibrium (at least 5000 time steps) of 50 simulation runs for each of the initial values, similarly to the approach used in [113, 151]. We use two ER networks with $N_{\text{A}}=N_{\text{B}}=100\,000$ nodes and average connectivity $\langle k_{\text{A}}\rangle=\langle k_{\text{B}}\rangle=20$ (as in [113]) and recovery rate $\alpha=0.002$, (as in [113, 151]).

4.5 Results

Fig. 4.1 shows bifurcation diagrams of the stationary disease prevalence I_{B}^{*} in network B as a function of the infection rate $\beta=\beta_{\text{A}}=\beta_{\text{B}}=\beta_{\text{inter}}$ for different values of the coupling probability

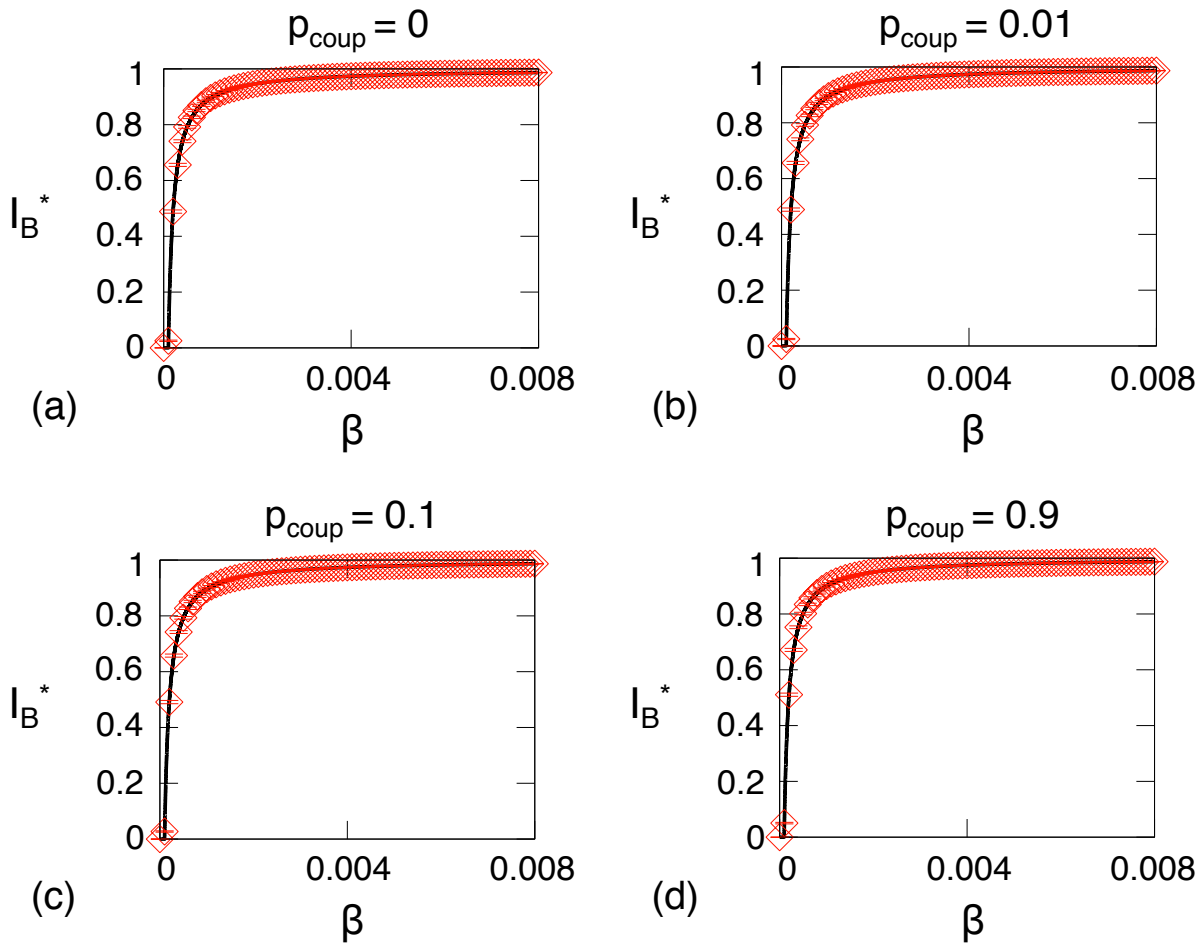


Figure 4.1: Bifurcation diagram of the stationary disease prevalence in network B, I_B^* as a function of the intra and inter-network infection rate, β when no rewiring is taking place for coupling probabilities (a) p_{coup}=0 (b) p_{coup}=0.01 (c) p_{coup}=0.1 (d) p_{coup}=0.9. Stable (thick black) branches have been computed analytically from Eqs. 4.1-4.9 using the dynamical systems analysis software XPPAUT [92]. Analytical results were confirmed using numerical simulation of two ER networks of 10^5 nodes and 10^6 edges (red diamonds). Points and error bars (if larger than marker size) were obtained by averaging at least 25 000 values corresponding to the prevalence at equilibrium (at least 5000 time steps) of 50 simulation runs for each of the initial values: ε_A=0.001; ε_B=0.001, ε_A=0.999; ε_B=0, ε_A=0.999; ε_B=0.999.

p_{coup} when no rewiring is taking place (γ=0). In this case, all coupling probabilities result in a single stable branch corresponding to continuous dynamical transition from healthy to endemic state which occurs at the epidemic threshold. The coupling has almost no effect in this case as the epidemic threshold is already very small in the case of a single network.

Fig. 4.2 shows the bifurcation diagrams obtained for rewiring rate γ=0.04. In the case where p_{coup}=0, our model recovers the result obtained in [112]: the two networks are separated and the resulting bifurcation diagram has two stable branches corresponding to lower and upper

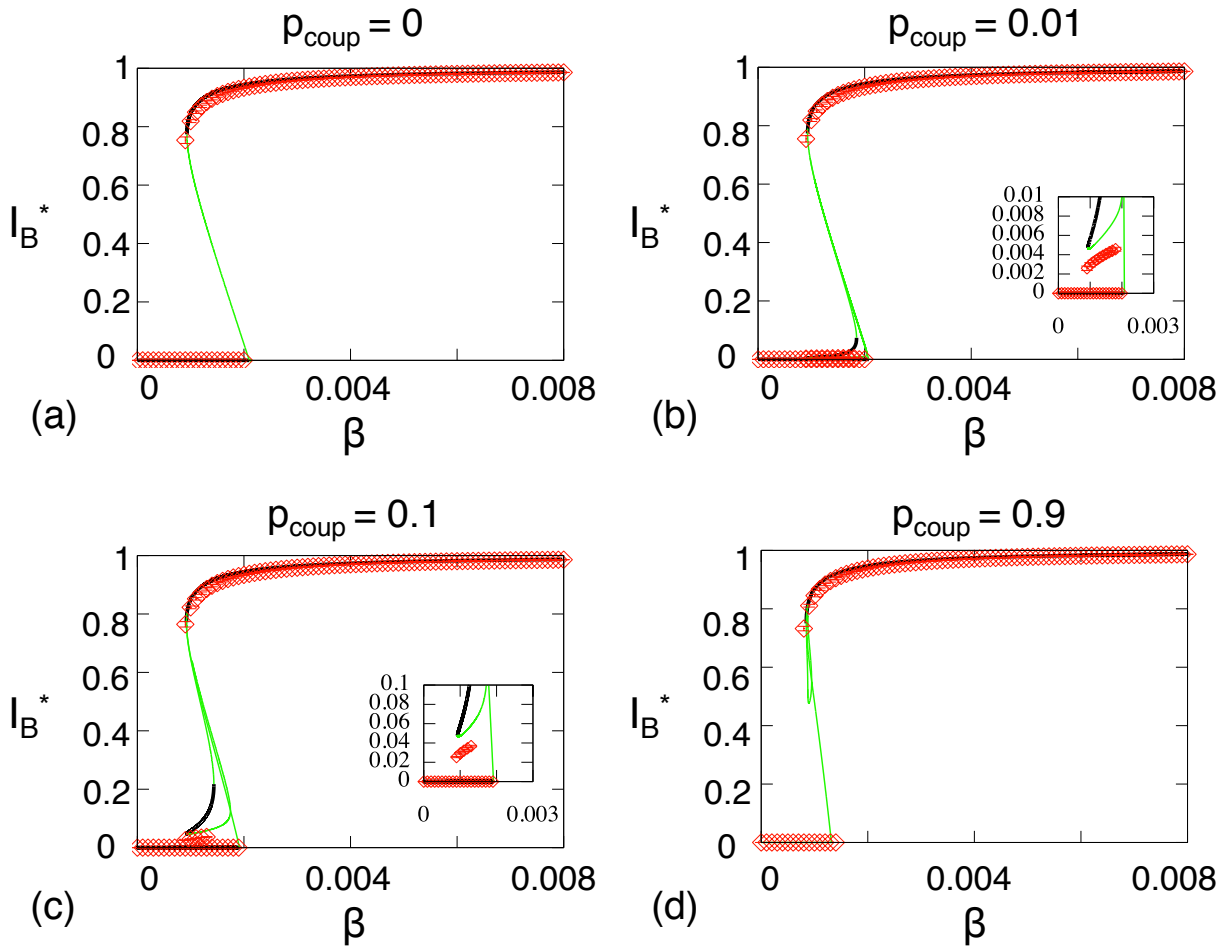


Figure 4.2: Bifurcation diagram of the stationary disease prevalence in network B, I_B^* as a function of the intra and inter-network infection rate, β for rewiring probability γ=0.04 and coupling probabilities (a) p_{coup}=0 (b) p_{coup}=0.01 (c) p_{coup}=0.1 (d) p_{coup}=0.9. Analytical computations of stable (thick black) and unstable (thin green) branches as well as numerical values (red diamonds) were obtained in the same way as in Fig. 4.1. The insets show a zoom of a new stable state in the cases of p_{coup}=0.01 and p_{coup}=0.1.

transition depending on the initial number of infected nodes. Numerical simulation is in good agreement with the results computed analytically.

In the cases of coupling probability p_{coup}=0.01, 0.1, both the results obtained analytically and numerically (shown more obviously in the inset) show a new stable state of intermediate I_B^* values, where the epidemic does not spread to all nodes in network B, but does not die out either. This state is obtained by starting the numerical simulation with no infected nodes in network B, ε_B=0, and almost all the nodes infected in network A, ε_A=0.999. In the following we show that this state describes the case where the epidemic does not spread equally to all the nodes in network B, but mostly to the coupled ones.

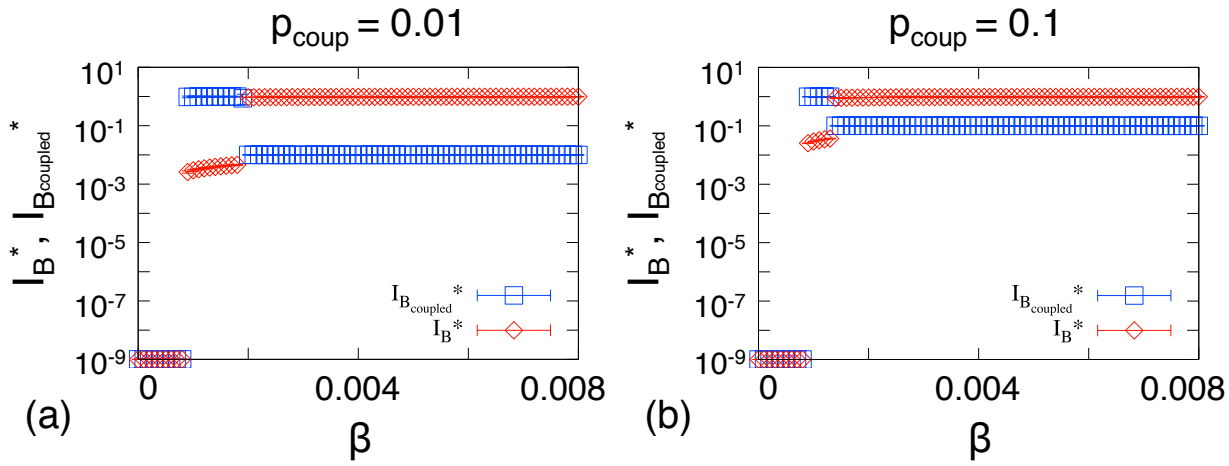


Figure 4.3: Fraction of infected nodes (red diamonds) and coupled infected nodes (blue squares) in network B at the end of the epidemic outbreak as a function of the intra and inter-network infection rate, β, for rewiring rate γ=0.04 and coupling probabilities (a) p_{coup}=0.01 (b) p_{coup}=0.1. Red diamonds show the fraction of infected nodes at the end of the epidemic outbreak, and the blue ones show how many of these nodes are coupled (see Eq. 4.11). Results obtained from numerical simulation of two ER networks of 10^5 nodes and 10^6 edges starting with initial values ε_A=0.999; ε_B=0. Mean and error bars (if larger than marker size) were obtained by averaging at least 25 000 values corresponding to the prevalence at equilibrium (at least 5000 time steps) of 50 simulation runs. The figures show that in the range where intermediate values of the disease prevalence are obtained, the fraction of coupled infected nodes is close to 1 and not to the coupling probability as expected. In other words, in this range mostly the coupled nodes are the ones who remain infected.

We define $I_{\text{B}_{\text{coupled}}}{}^*$, the fraction of coupled nodes in network B that remain infected at the end of the epidemic outbreak, as

$$I_{\text{B}_{\text{coupled}}}{}^* = \begin{cases} \frac{N_{\text{B}_{\text{coupled infected}}}{}^*}{N_{\text{B}_{\text{infected}}}{}^*}, & \text{if } N_{\text{B}_{\text{infected}}}{}^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.11}$$

where $N_{\text{B}_{\text{infected}}}{}^*$ and $N_{\text{B}_{\text{coupled infected}}}{}^*$ are respectively the number of infected and coupled infected nodes in network B at the end of the epidemic outbreak.

Fig. 4.3 shows $I_{\text{B}_{\text{coupled}}}{}^*$ and I_{B}^* as a function of the infection rate β when starting with initial conditions ε_A=0.999 and ε_B=0. In a network where the infected nodes are equally distributed, we would expect the fraction of infected coupled nodes $I_{\text{B}_{\text{coupled}}}{}^*$ to constitute the same part of the infected nodes as the fraction of coupled nodes of all of the nodes. In other words, we would expect $I_{\text{B}_{\text{coupled}}}{}^*$ to be close to p_{coup} (unless there are no infected nodes, where it would be 0). Indeed, the graph starts with y values close to 0, indicating that there are no infected nodes. But, in the range of β where intermediate I_{B}^* values are obtained ($0.0008<\beta<0.002$ for p_{coup}=0.01

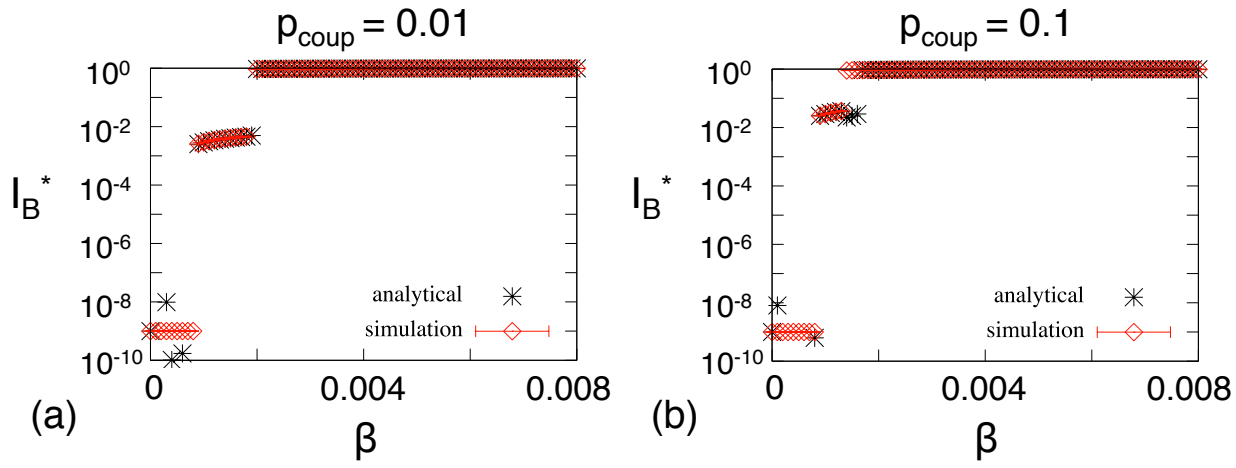


Figure 4.4: Fraction of infected nodes in network B at the end of the epidemic outbreak as a function of the intra and inter-network infection rate, β, for rewiring rate γ=0.04 and coupling probabilities (a) p_{coup}=0.01 (b) p_{coup}=0.1 starting with initial conditions ε_{A}=0.999; ε_{B}=0. Red diamonds correspond to results obtained from numerical simulation of two ER networks of 10^5 nodes and 10^6 edges (same as in Fig. 4.2). Mean and error bars (if larger than marker size) were obtained by averaging at least 25 000 values corresponding to the prevalence at equilibrium (at least 5000 time steps) of 50 simulation runs. Black points correspond to results obtained analytically from Eqs. 4.1-4.9, where pair approximation of the triplets $[I_{\text{A}}S_{\text{A}}I_{\text{B}}], [S_{\text{A}}S_{\text{A}}I_{\text{B}}], [I_{\text{A}}S_{\text{A}}S_{\text{B}}]$ is substituted with the mean values obtained from the simulation.

and $0.0008<\beta<0.0014$ for p_{coup}=0.1), the fraction of infected coupled nodes ${I_{\text{B}_{\text{coupled}}}}^*$ is close to one, indicating that almost all infected nodes are coupled. For larger β values, the expected result is obtained and ${I_{\text{B}_{\text{coupled}}}}^*$ is close to the coupling probability p_{coup}.

This explains the quantitative disagreement between the analytical solution and the numerical simulation. Recall that the pair approximation assumes homogeneous distribution of links in the network. Therefore, just as the approximation does not hold during the initial phases of the dynamics, in a partially invaded network (as shown in Fig. 4.3) correlations exist that are not captured by the approximation. The emergence of this new stable state, observed both numerically and analytically for the case of coupling probability p_{coup}=0.01, 0.1, is a consequence of the fact that the coupled nodes can not rewire their inter-network links. Therefore, for low infection rates, the disease can be established in one network while not invading the other but instead reaching only to the coupled nodes, since the nodes that are not coupled can protect themselves by rewiring their links. This is not possible for larger infection rates where the new stable state is no longer reachable and only the two other states of endemic and disease-free remain stable.

In the following we confirm the deviation of the simulation from the analytical solution (shown in Fig. 4.2) originates from the pair approximation of the triplets $[I_{\text{A}}S_{\text{A}}I_{\text{B}}], [S_{\text{A}}S_{\text{A}}I_{\text{B}}], [I_{\text{A}}S_{\text{A}}S_{\text{B}}]$. We do this by substituting the pair approximation of these triplets with the actual densities obtained

from the simulation, and leave everything else in Eqs. 4.1-4.9 the same.

Fig. 4.4 shows both numerical and analytical results for γ=0.04 and p_{coup}=0.01, 0.1 starting with initial conditions ε_{A}=0.999 and ε_{B}=0. Following the observation above that the assumption of homogeneous distribution of links in the network does not hold in this case (since the epidemic persists only in the coupled nodes) we substitute the pair approximation of all the triplets involving nodes from both networks with the values obtained numerically. The graph shows a very good agreement between results obtained numerically and analytically, indicating that the deviation in Fig. 4.2 was indeed as a result of the pair approximation for these triplets. As mentioned in section 4.3, there are other methods to overcome the inadequacy of the pair approximation in the presence of long range correlations, such as higher order approximation [151] or using epidemic threshold calculation [43, 112]. This is not in the scope of the current work but can be addressed in future more accurate models.

The third stable state shown in Fig. 4.2 for the cases of coupling probability p_{coup}=0.01, 0.1, becomes unstable again for strongly coupled networks (p_{coup}=0.9). This result is expected since in tightly coupled networks, there exist a larger number of coupled nodes that can not protect themselves from infection due to inter-network links even at small infection rates. Therefore, any epidemic persisting in one network, will eventually spread and persist in the other network as well.

Finally, we see in Fig. 4.2 that the width of the multistability area becomes smaller as the coupling probability increases. In other words, the range of β values for which more than one stable state is reachable depending on the initial conditions becomes smaller with more inter-network links meaning that a system of tightly coupled networks is more stable and depends less on the initial conditions.

4.6 Conclusion

In this chapter, we have developed an analytical formalism to study epidemic spreading in coupled adaptive networks, where intra-network links are rewired based on the states of the nodes. We have demonstrated that inter-network links between otherwise independently adaptive networks can result in qualitative changes in the networks dynamics, and thus should be taken into consideration in future models. In particular, we find a new stable state (that does not exist in the case of a single isolated adaptive network) where in one network the epidemic persists throughout the network and in the other it only persists in the coupled nodes and does not break-out into the rest of the network. We also show that a system of coupled adaptive networks is more stable (less dependent on initial conditions) with the existence of more inter-network

links.

Our analytical formalism is flexible allowing each network to have different intranetwork infection rates as well as different starting configurations, and it also allows different infection rates between (inter-) and within (intra-) the networks, accounting for limited cases of time-scale separation between the intra- and internetwork disease-spreading processes. However, even in the simple case of homogeneous ER networks, one must account for the actual second-order moments of the system in order to obtain the exact quantitative behaviour of the new observed equilibrium. In section 4.3 we discussed possible approaches to extend the model in order to describe the behaviour of more complex systems, such as coupled scale-free networks. But although limited, our model provides a starting point for future more complicated models and hopefully opens new avenues in the study of interconnected adaptive networks, which represent most real-world systems [39].

Our current study already raises some interesting future questions. For example, how does rewiring of inter-network links affect the dynamics? In the example mentioned earlier of coupled nodes representing people who travel a lot, rewiring of inter-network links corresponds to the imposition of travel restrictions. In this case, we expect the new stable state discovered to disappear for sufficiently high rewiring rates of the inter-network links. The effect of time scale separation between the epidemic spreading processes taking place within and between the networks is also a very important point. Finally, it would be very interesting to study a system consisting of any number of coupled adaptive networks within this model, and not just two as in the current work, and see if more networks will lead to more stable states, as happened with the extension from one to two networks. We suspect that there might be more stable states, for example, in the case where some nodes are coupled with more than one network (i.e. there is an overlapping between the groups of nodes that are chosen to be coupled with each network). In this case, there might be a stable state for each group of nodes that are coupled with more than a certain number of networks, thus revealing dependency groups (nodes tightly coupled) in the system.

CHAPTER FIVE

CONSTRAINED EPIDEMIC SPREADING ON CORRELATED COUPLED NETWORKS

Positive correlations between the local properties of interconnected nodes, e.g. intra-network degree, in a system of coupled networks have been suggested to increase the connectivity of the overall system, thus enhancing the spreading of flows. Indeed, many real-world networks, including social, transportation and financial networks, have been found to display strong positive correlations, however, the random correlations found in coupled biological networks still remain a mystery. Inspired by this observation, in the current chapter we demonstrate that in the presence of resource constraints, as is often the case in coupled networks, positive correlation can impede flow processes through contention, thus providing a possible explanation for the random coupling observed in biological networks, and illustrating the importance of incorporating more realistic scenarios in future models for coupled complex networks. The chapter is based on the author's work [203].

5.1 Introduction

Coupled networks are often formed by positively correlated networks, resulting in nonrandom but structured coupling called *intersimilarity* [179], where nodes are coupled according to some regularity. In particular, in many real-world systems, the degree of a node is positively correlated

with the degrees of its coupled nodes. For example, Parshani *et al.* [179] studied a system composed of an interdependent world wide port network and a world wide airport network and showed that well-connected ports tend to couple with well-connected airports. The international trade multiplex networks, where countries are connected by various commodity trades, has been shown to display strong positive correlations between the number of connections a country has in each commodity-specific layer [25]. Finally, the degrees of nodes in different layers of interactions in a multirelational social network extracted from a massive multiplayer online game, were found to be highly correlated, i.e. players who communicate with many others via one type of social interaction are likely to communicate with many others also via other types of social interactions [224].

Motivated by these empirical findings, recent studies examine the effect of intersimilarity on the behaviour of coupled networks. Parshani *et al.* [179] showed that positive correlations between the degrees of interdependent nodes increase the robustness of a system of two interdependent networks to random failure. Buldyrev *et al.* [54] showed that when mutually dependent nodes have the same number of intra-network connectivity links, the resulting percolation threshold is *always* smaller than the one for randomly interdependent networks with the same degree distribution. Positively correlated multiplex networks, where nodes' degrees are positively correlated across layers, have been shown to promote the emergence of the giant component [64, 145].

The studies mentioned above suggest that intersimilarity between coupled networks increases the connectivity of the overall system, thus enhancing the spreading of flows in the system (a result that we confirm here, and was also confirmed later by Zhao *et al.* [248]). However, the random coupling observed in several biological networks suggests that there might be another competing mechanism, leading to a trade-off between negative and positive correlations. The multiplex cellular network of a bacterium called Mycoplasma pneumoniae, in which proteins are connected through both physical bindings and by sharing metabolites, has been found to behave similarly to its randomized version (obtained by randomly shuffling the node identities in the two layers) [64]. This is in contrast to a world-trade network, studied in the same paper, which was found to behave similarly to its maximally positively correlated version (obtained by ordered-matching of degree ranks in the two layers). Pocock *et al.* [188] studied a multiplex ecological network of multiple species' interactions (e.g. flower–flower visitor network, flower–butterfly network) and found no significant correlations between the importance of nodes in different layers.

Indeed, we note that unlike connected nodes in a single network, coupled nodes often share limited resources, and this will clearly affect any flow process operating over the network. This is obviously the case where the coupled nodes represent the *same* entity, as in a multiplex,

but it is also true in the case where the coupled nodes represent *different* entities that interact with or depend on each other. The coupled nodes play a "role" in more than one network, and typically have limited resources available. For example, in the case where an overlay node in a communication network is coupled with a router through a physical device (i.e. a host computer) the router's packet queue handles both the packets that are routed to the host and those in the network layer that do not intersect with the host. Counter-intuitively it might be better to couple a central overlay node with a non-central router, so that the router is "dedicated" to the overlay node and is not being heavily loaded with routing general network packets. Similarly, in the case where a person is involved in more than one social network through different types of social links, the person shares his/her time between contacts in different social groups and thus might not be able to play a central role in spreading (for example) news equally amongst all the different social groups.

We consider a model of constrained epidemic spreading in a multiplex network, where nodes are constrained to interact with (and therefore potentially infect) a maximum number of their neighbours at each epidemic time step. This is equivalent to the case whereby a person can only meet a certain number of his/her friends in a period of time: it seems clear that in the case of frequency-dependent processes (such as for sexually-transmitted diseases [228]) such contact limitations will potentially affect the spread of the disease through the network. Our results show that, in the absence of constraints (i.e. "regular" epidemic spreading), positive correlation results in a smaller epidemic threshold than negative, in agreement with previous studies showing that positive correlation increases the robustness of network, and thus results in better connectivity [64, 145, 179]. This result was also confirmed later by another study [248]. However, in the presence of constraints, the result is qualitatively different, where the epidemic threshold obtained through positive correlation is larger than the one obtained through negative correlation: intersimilar coupled networks spread less efficiently than negatively correlated networks in the presence of constraints. As a consequence of these results we suggest that future work must account for real-world limitations when considering coupled networks.

The rest of this chapter is organised as follows: in section 5.2 we present our model of constrained epidemic spreading over a correlated multiplex, and in section 5.3 we describe the computer simulation used to numerically study the suggested model. In section 5.4 we analyse the obtained results and provide a mathematical explanation based on the degree distribution and the degree-degree correlations that emerge in the multiplex network as a result of the constraints. We finally summarising our findings in section 5.5.

5.2 Model

The structure of the coupled network we consider is a multiplex consisting of two networks sharing the same set of N nodes, also called *duplex*. For simplicity and clarity of the results, we consider only the two extreme cases of maximally positive or maximally negative correlations. Imperfect correlation can be considered, for example, by coupling a fraction of the nodes in a maximally correlated manner, and the rest in an uncorrected manner, as done in [145]. In order to construct a maximally positively (negatively)-correlated duplex, we sort the nodes according to their degree in each network, and match two nodes from each network in order (in opposite order) of the degree rank [64, 145]. The degree of a node in the obtained multiplex is the sum of degrees in each network minus the number of overlapping links in the two networks, which can be neglected in the limit $N \to \infty$ for random, sparse networks with the largest degree of at most order $\sqrt{(}N)$ [167]. Therefore, the way in which the two initial network layers are overlaid one another to construct the interlaced network, could affect the connectivity of the obtained duplex. In particular, correlated coupling, as considered here, can introduce higher-order correlations, such as degree-degree correlations, corresponding to the correlations between the degrees of connected nodes in the duplex. But generally, correlated coupling does not necessarily imply degree-degree correlations in the obtained system of coupled networks, for example, in the case of maximally correlated coupled ER networks with the same degree distribution, the degree-degree correlation in the obtained duplex is close to zero [145]. However, this is an important issue, on which we expand in section 5.4. Finally, in order to distinguish between the effects of constraints and correlated coupling, we also consider an uncorrelated duplex where nodes are randomly coupled.

The dynamics that we consider is a constrained SIR epidemic spreading. Recall that in the well-established SIR model described in chapter 2, each node exists in one of three discrete states: susceptible (S), infected (I), or recovered (R). Starting with a network of susceptible nodes and introducing a seed of infected nodes, at each time step, each susceptible (healthy) node is infected with infection rate λ if it is connected to one or more infected nodes in the duplex. At the same time, infected nodes recover with recovery rate μ, after which they are immune to further infection. In our model of constrained epidemic spreading, nodes are limited to interact with a maximum number of their neighbors m (called *interaction limit*) at each epidemic time step. In order to apply this limit, at each time step, we extract an *interaction graph* from the original duplex network, called the *original graph*, by successively selecting a node at random from the original graph that has not yet reached its interaction limit, and randomly choosing one of this node's neighbors that has also not reached its own interaction limit. We continue until no interactions can be added. To prevent the first nodes from exhausting their interaction limit

early we shuffle the node order before each selection. The choice of neighbors with which to interact is made independently of their epidemic state S, I or R. Given this *interaction graph*, we update the state of each node according to the SIR rules: susceptible nodes with infected neighbors becoming infected with a probability λ, and infected nodes recovering at a rate μ. Note that if we were to pick nodes below the limit and then apply the SIR rules immediately, we might double-up on applying rules twice to one node. Moreover, we later analyse the properties of the resulting interaction graphs in order to understand the implications of the interaction limit, and differentiate them from the implications of correlated coupling.

5.3 Simulation

We consider a duplex consisting of either two ER or two BA networks of size $N = 100\,000$ and average connectivity $\langle k_1 \rangle = \langle k_2 \rangle = 6$, similar parameters to the ones used in other epidemic spreading studies [183, 184]. We first consider interaction limits m ranging from 10 to 14, which is two links less or more than the approximate mean degree of the duplex, $\langle k \rangle = 12$. Later, we also examine how changing the interaction limit from 0 to 100, while keeping the infection rate constant, affects the results. Without loss of generality we set the recovery rate to unity, $\mu = 1$, as different recovery rates can be considered by a proper re-scaling of λ, m and time. Note that at this point we assume that the interaction limit and the disease are taking place on the same timescale, i.e. both are carried out using the same discrete time steps of length Δt. The more general case of separate timescales might also be addressed by an appropriate re-scaling of m. For example, in case of a disease evolving over months though a population limited to meeting a certain number of friends in a week, the interaction limit can be re-scaled in order to obtain the number of friends a person is limited to meet in a month. However this re-scaling is not always possible, and depends on the nature of the dynamics. We performed 1000 simulation runs, our results obtained by averaging over 100 random starting configurations, on 10 different realizations of the random networks (a common approach in simulation of epidemic spreading on networks [183, 184, 186]).

In order to obtain reproducible results, and since the construction of the interaction graph is computationally expensive, we first create a collection of interactions graphs for each network seed and epidemic time step. Thus, for each network parameters and a given network seed, we create, in a loop, 500 interactions graphs (which we assume as the number of epidemic time steps required), where each iteration is initialising the pseudorandom number generator based on the given network seed and the current time step. Note that 500 time steps is generally more than enough for the dynamics to halt (with all nodes either recovered or susceptible) under the synchronous updating scheme (where nodes states are updated in parallel at each time step), as

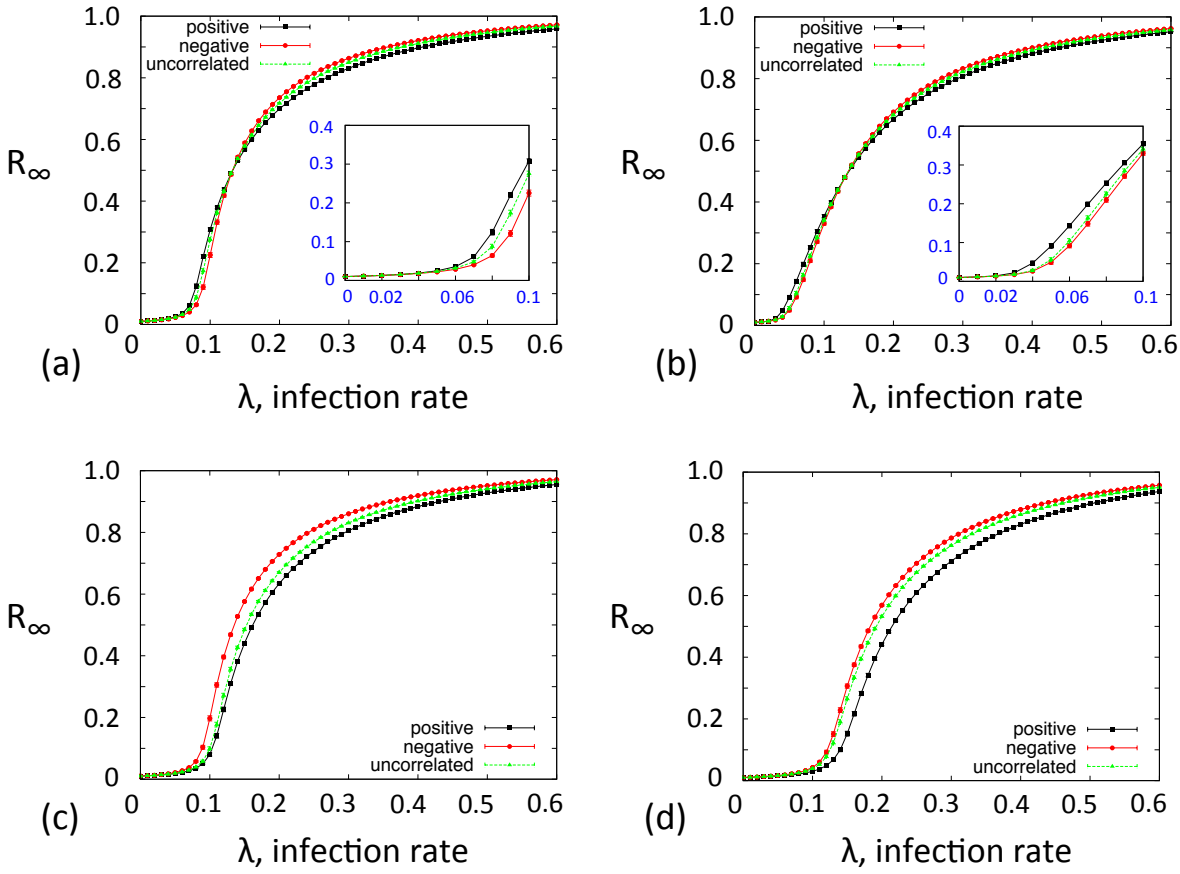


Figure 5.1: Fraction of infected nodes at the end of the epidemic outbreaks R_∞ as a function of the infection rate λ for constrained and non-constrained epidemic spreading, running on a positively-correlated (black squares), negatively-correlated (red circles) and uncorrelated (green triangles) networks consisting of two ER or BA networks of size $N = 100\,000$ and mean degree 6. Points and error bars (if larger than marker size) correspond to the mean and standard deviation computed over 10000 simulation runs. (a)-(b) non-constrained epidemic on coupled ER and BA networks respectively. (c)-(d) constrained epidemic with interaction limit m=12 on coupled ER and BA networks respectively.

the epidemic develops exponentially in homogeneous networks and even faster in heterogeneous ones [27, 163]. If, however, infected nodes still remain in the network, we continue by using the previous interactions graphs created (using the current time step modulo 500).

5.4 Results

In Fig. 5.1 we show the fraction of recovered nodes at the end of an epidemic outbreak R_∞ as a function of the infection rate λ in the cases of constrained and regular epidemic spreading, starting with a fraction of 0.01 infected nodes. Points and error bars (where shown) correspond to the mean and standard deviation computed over 1000 simulation runs; where error bars are

not shown, they would be smaller than the marker size. In the cases of regular non-constrained epidemic spreading, see Fig. 1(a)-(b), the curve for positive correlation (black squares) is going up earlier than the curve for negative correlation (red circles) for both ER and BA networks. In other words, the epidemic threshold of positively-correlated coupled networks is smaller than that for the negatively-correlated case. If we define the epidemic threshold as the largest infection rate in which $R_\infty < 0.1$, then we obtain from the simulation that the epidemic threshold for positively-correlated coupled networks is $\lambda_c \approx 0.07$ for ER and $\lambda_c \approx 0.05$ for BA, while that for negatively-correlated networks is $\lambda_c \approx 0.08$ for ER and $\lambda_c \approx 0.06$ for BA.

However, in the case of constrained epidemic spreading with interaction limit $m = 12$, see Fig. 1(c)-(d), both positively-correlated ER and BA networks exhibit larger epidemic thresholds ($\lambda_c \approx 0.1$ and $\lambda_c \approx 0.14$ respectively) than in the negatively-correlated cases ($\lambda_c \approx 0.08$ and $\lambda_c \approx 0.12$ respectively). Finally, in the case of regular epidemic spreading, once the infection rate is larger than the threshold, it grows more abruptly in the negative-correlation case than in the positive (the red curve crosses the black curve and stays above it). This result agrees with earlier work [64, 145] showing that the giant component in a positively-correlated multiplex emerges at lower link densities than for the negatively-correlated multiplex but, once formed, grows much more gradually.

As expected, in all 4 figures above, the green curve, representing uncorrelated networks, is between the red and black curves. However, uncorrelated coupled ER networks behave more similarly to positively-correlated coupled ER networks, where for BA networks it is the opposite. This is as a result of the obtained degree distributions when coupling ER or BA networks. In the following we examine the degree distributions of the original and interaction graphs obtained from the simulation. Recall that the original graph is the duplex consisting of two random networks, and interactions take place by constraining the maximum number of neighbors in the original graph with which a node can interact at each time step. Fig. 5.2 shows the degree distributions obtained by averaging over all the original and interaction graphs. The original graph of positively-correlated networks display a zigzag pattern (see Fig. 2(a)-(b)) originating in the fact that for large enough networks ($N \to \infty$), each node has exactly the same degree in both networks [145]. The original graph consisting of uncorrelated coupled ER networks behaves more similarly to the positively-correlated original graph than the negatively-correlated one. This is due to the narrow degree distribution of a single ER network, where most of the nodes have degree close to the mean (6 in this case). Therefore, when coupling the nodes in a negatively-correlated manner, there is a very large peak at the mean degree of the duplex (12 in this case). However, uncorrelated or positively-correlated coupling yields a higher chance to degrees further away from the mean to be obtained resulting in a broader degree distribution. In

the case of BA networks, the opposite process is happening: as the degree distribution of a single BA network is broad, positively-correlated coupling gives rise to very high degrees that have very small chance to be obtained in the case of negatively-correlated and uncorrelated coupling. By introducing a constraint on a network, we narrow its degree distribution, reducing the high fluctuations in connectivity. Therefore, networks with narrow degree distribution in the first place (as the negatively-correlated original graph) are less affected by this process.

Recall that the mean field approximation for the epidemic threshold in the thermodynamic limit $N \to \infty$ is given by

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{5.1}$$

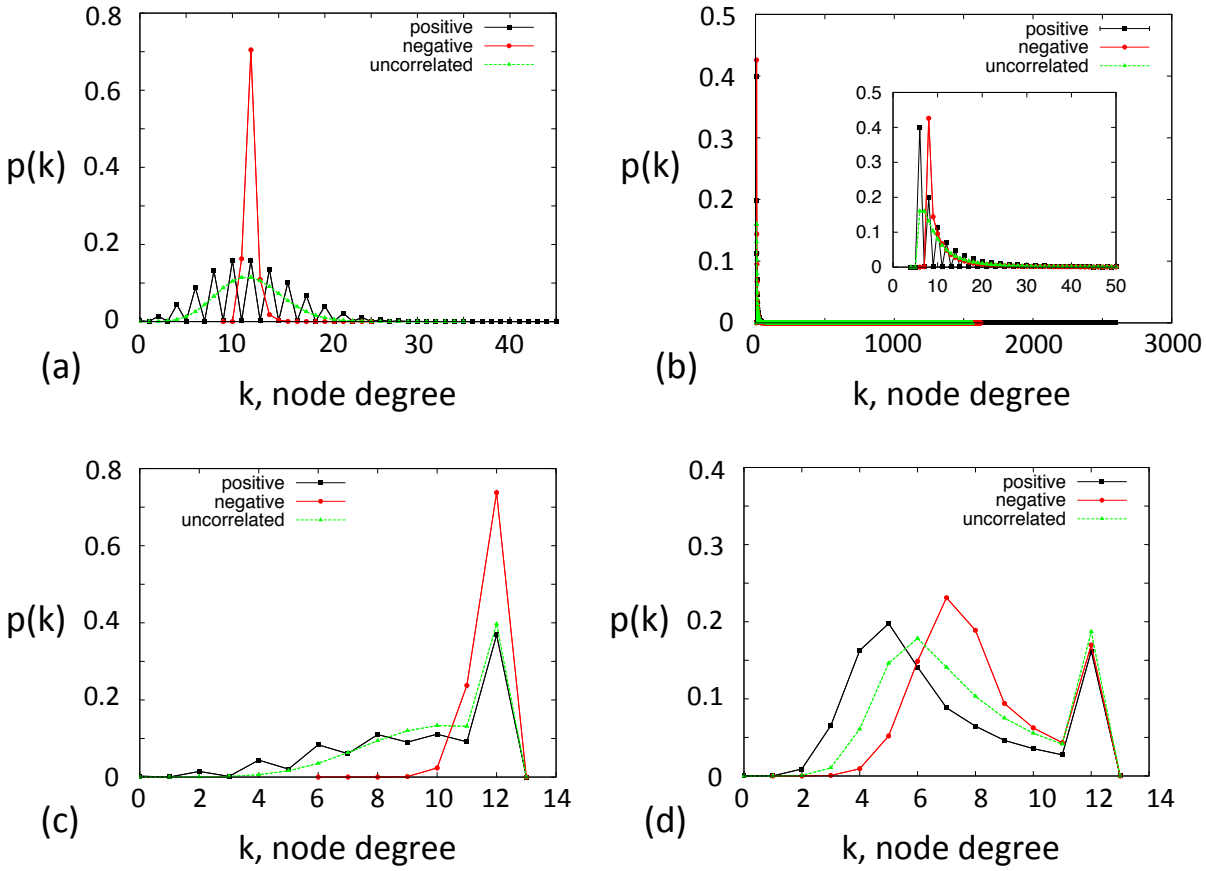


Figure 5.2: Degree distributions of the original and the interaction graphs, obtained from a positively-correlated (black squares), negatively-correlated (red circles) and uncorrelated (green triangles) networks consisting of two ER or BA networks of size $N = 100$ and mean degree 6. Results obtained by averaging all the graphs. (a)-(b) original graphs consisting of two ER networks and two BA networks respectively. (c)-(d) interactions graphs with interaction limit $m = 12$ obtained from coupled ER and BA networks respectively.

where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments of the network degree distribution respectively [163]. This result is also demonstrated in chapter 2 where we derive the percolation threshold, which can be mapped exactly to the SIR epidemic threshold [111]. The epidemic threshold is the critical infection rate in which larger infection rates $\lambda > \lambda_c$ result in a finite fraction of infected nodes, and smaller infection rates $\lambda < \lambda_c$ result in an infinitesimally small number of infected nodes in the limit of very large networks. Networks with strongly-fluctuating connectivity distributions, like scale-free networks, show a vanishing epidemic threshold for increasing network sizes, $\langle k^2 \rangle \to \infty$ for $N \to \infty$.

Thus, by applying a constraint on a positively-correlated network, we decrease its degree fluctuations i.e. the second moment, which is the denominator in Eq 5.1. But since positively-correlated networks have more high-degree nodes, the constraints also result in a decrease in the mean degree (the numerator in Eq 5.1). However, the decrease in the second moment is more significant, resulting in original moments for the ER networks of $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{12}{168.08} = 0.071$ for positively-correlated network and $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{12}{144.46} = 0.083$ for negatively-correlated. After applying the constraint, the new values are $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{9.51}{97.21} = 0.098$ for positively-correlated network and $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{11.82}{139.94} = 0.084$ for negatively-correlated.

In both constrained and regular epidemic spreading on a duplex of ER networks, the mean field gives a good approximation for the values obtained from the numerical simulation yielding epidemic threshold of $\lambda_c \approx 0.07$, $\lambda_c \approx 0.08$ for regular epidemic spreading, and of $\lambda_c \approx 0.1$, $\lambda_c \approx 0.08$ for constrained epidemic spreading on positively- and negatively-correlated networks respectively. Also, the values obtained from the simulation and the approximation both show a small difference between constrained and non-constrained epidemic spreading on negatively correlated coupled ER networks. The reason is that most of the nodes in a duplex of negatively-correlated ER networks have degree equal to or smaller than 12, which is the interaction limit. Therefore, the constraint does not affect most of the nodes, and a similar epidemic threshold is obtained.

In the case of regular epidemic spreading on a duplex of BA networks, although qualitatively in agreement, the values obtained from simulation $\lambda_c \approx 0.05$, $\lambda_c \approx 0.06$ deviate from those obtained by Eq. 5.1 $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{11.99}{564.26} = 0.021$, $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{11.99}{339.45} = 0.035$ for positively- and negatively-correlated networks respectively. In the case of constrained epidemic spreading on coupled BA networks, the values obtained from the mean field approximation are $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{6.85}{55.37} = 0.123$ and $\frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{8.3}{73.53} = 0.11$ for the positively- and negatively-correlated networks respectively. These values are closer (especially for the negative correlation) to the values obtained from the simulation: $\lambda_c \approx 0.14$ and $\lambda_c \approx 0.12$ for positively- and negatively-correlated networks

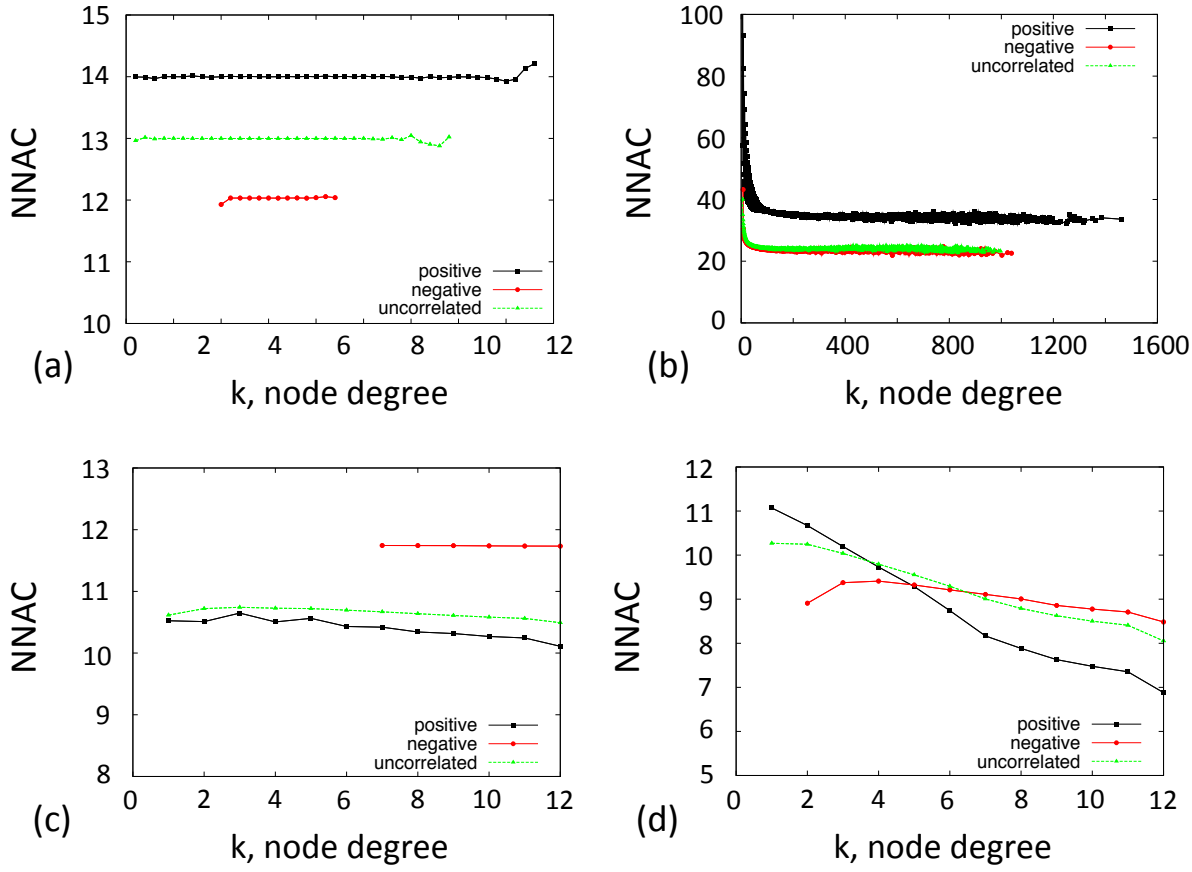


Figure 5.3: Nearest neighbors average connectivity (NNAC) as a function of node degree for a positively-correlated (black squares), negatively-correlated (red circles) and uncorrelated (green triangles) networks consisting of two ER or BA networks of size $N = 100$ and mean degree 6. Results obtained by averaging all the graphs.(a)-(b) original graphs consisting of two ER networks and two BA networks respectively. (c)-(d) interactions graphs with interaction limit $m = 12$ obtained from coupled ER and BA networks respectively.

respectively.

In order to explain these differences, we examine the effect of degree correlated coupling and interaction limit on the degree-degree correlations of the obtained original and interaction graph, which are not taken into consideration in the mean field approximation. As mentioned before, degree-degree correlation is the tendency of nodes to *connect* with similar degree nodes, where degree correlated coupling is the tendency of nodes to *couple* with similar degree nodes. We use both the nearest- neighbors average connectivity (NNAC) curve [182] and the Pearson correlation coefficient of the degrees of nodes at either end of a link in the duplex [171] as a measure for degree-degree correlation. Figures 5.3(a) and 5.3(b) show the NNAC curves obtained for the original graphs (only correlated or uncorrelated coupling, no interaction limit)

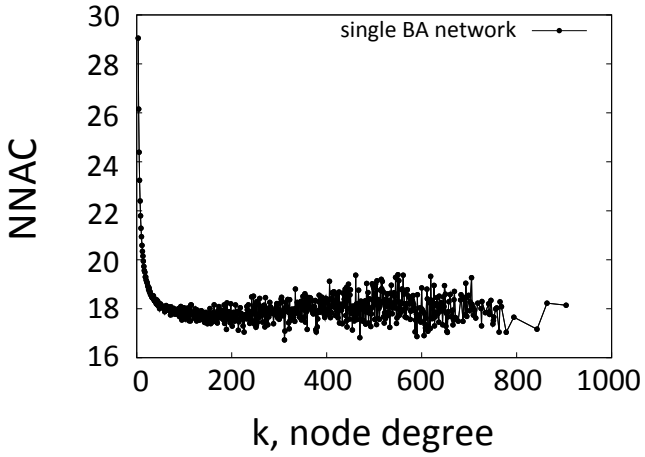


Figure 5.4: Nearest neighbors average connectivity (NNAC) as a function of node degree for a single BA network of size $N = 100\,000$ and mean degree 6. Results obtained by averaging 1000 networks.

consisting of ER and BA networks respectively. In the case of coupled ER networks, see Fig. 5.3(a)), all 3 curves show almost no degree-degree correlations. This is also confirmed by the obtained Pearson correlation coefficients: $-5.288 \times 10^{-5} \pm 0.0013$, $4.346 \times 10^{-5} \pm 0.0012$ and $-2.236 \times 10^{-5} \pm 0.0012$ for positively-correlated, negatively-correlated and uncorrelated ER networks respectively. This result is in agreement with [145], showing that the positive degree-degree correlations of a multiplex network consisting of two correlated coupled ER networks vanishes at $\langle k_1 \rangle = \langle k_2 \rangle$, which is exactly what happens in our simulation.

The non-constrained original graph consisting of two BA networks display very similar degree-degree correlation to the one obtained for the single BA network (see Fig. 5.4): the NNAC curve shows power-law dependence, in agreement with [232], while the Pearson correlation coefficient (-0.02 ± 0.001, -0.013 ± 0.001, -0.017 ± 0.0014 and -0.018 ± 0.002 for positively-correlated, negatively-correlated and uncorrected duplex and a single BA network respectively) is close to zero, in agreement with [171]. The power-law behavior of the NNAC curve might explain the deviation from the mean field approximation, as the approximation neglects the connectivity correlations in the network.

Figures 5.3(c) and 5.3(d) show the NNAC curves obtained for the interaction graphs constrained by interaction limit $m = 12$ consisting of ER and BA networks respectively. Pearson correlation coefficients obtained are -0.055 ± 0.001 and -0.34 ± 0.002 in case of positively-correlated ER and BA networks respectively, and -0.002 ± 0.001 and -0.12 ± 0.001 in case of negatively-correlated ER and BA networks respectively. In both the case of ER and BA networks, the positively-correlated graphs exhibit stronger negative degree-degree correlations than negatively-

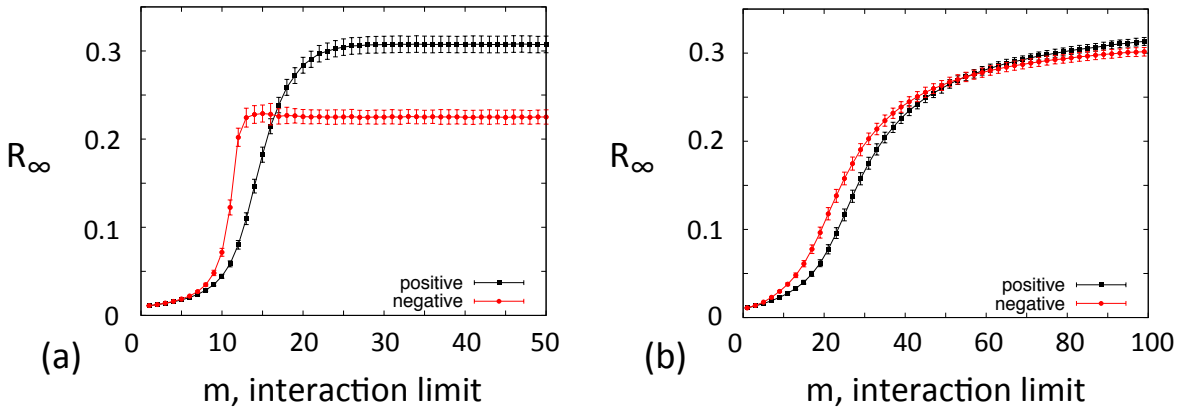


Figure 5.5: Fraction of infected nodes at the end of the epidemic outbreaks R_∞ as a function of the interaction limit m for constrained epidemic spreading with infection rate $\lambda = 0.1$, running on a positively (black squares) and negatively (red circles) correlated networks consisting of two ER (a) or BA (b) networks of size $N = 100\,000$ and mean degree 6. Points and error bars (if larger than marker size) correspond to the mean and standard deviation computed over 1000 simulation runs.

correlated graphs. This is due to the construction of the interaction graph which creates a tendency of high-degree nodes to have lower-degree nodes as neighbors and has a stronger effect on broad degree distribution: In order to have a degree closer to the interaction limit, a node needs its neighbors not to reach their limit before it chooses to interact with them. As lower-degree nodes have fewer neighbors with which to interact, they "wait" to interact with the node. By analogy, a person whose friends are high-degree nodes might want to meet them at the end of the week, but might find them are too tired since they have met many friends during the week (they have reached the limit). In contrast, his/her low-degree friends will be happy to meet him/her since they have not yet exhausted their limits. This process produces stronger degree-degree correlation in cases where the original graph has broad degree distribution, as in the case of positively-correlated BA networks. And indeed, the deviation of the simulation values from the mean-field approximation is more signifiant in the case of constrained epidemic spreading on positively-correlated BA networks (0.14 vs 0.123), than in the case of negatively-correlated BA networks (0.12 vs 0.11). The uncorrelated case, as in previous figures, behaves more similarly to the positively-correlated case for ER networks and the opposite for BA networks.

Finally, we examine how changing the interaction limit m affects the results. Until now we have used an interaction limit of $m = 12$. In Fig. 5.5, we show R_∞, the fraction of recovered nodes at the end of the epidemic outbreaks, as a function of m where $\lambda = 0.1$. Both graphs start with the red curve (corresponds to negative coupling) above the black curve (positive coupling), which means that negatively-correlated coupled ER and BA networks result in a larger number of infected nodes using the same interaction limit and the same infection rate. At some point,

$m = 17$ in case of ER and $m = 55$ in case of BA, the curves cross each other, and afterwards the positively-correlated networks become more efficient in spreading the epidemics. This means that the interaction limit is no longer impeding the spreading of the epidemic: it is sufficiently high not to function as a limit *per se*, and does not limit nodes' interactions in the network.

5.5 Conclusion

In this chapter we have studied the behaviour of multiplex networks obtained by correlated coupling of two random networks, which received much focus in recent studies. We propose a model of constrained SIR epidemic spreading capturing a common situation in coupled networks, and in multiplex networks in particular, where coupled nodes share limited resources such as time, energy, and memory. First we confirm that "regular" non-constrained epidemic is more easily spread in positively correlated multiplex, as suggested by previous studies showing that positive correlation increases the connectivity of the overall network. However, in contrast with recent studies, positively-correlated coupled networks are less efficient in spreading the epidemic than negatively-correlated networks in the presence of constraints. By obtaining a qualitatively different result when considering constraints, we illustrate the importance of incorporating more realistic scenarios in future models for coupled complex networks. In particular, when extending models from single to multiple networks, such as percolation or epidemic spreading, it is of great importance to carefully consider the qualitative differences between single isolated networks and their coupled interacting counterparts.

The effect of correlations and nonrandom coupling on the behaviour of coupled networks is still a very active research topic. More recently, researchers have been examining the formation of overlapping links, where two friends, for example, communicate both via email and via mobile phone, or in transportation networks two cities connected by a main road are likely to be connected also by a railway [63]. Links overlap has also been found to increase the robustness of the whole system [63, 128], and various coupled networks formation models demonstrate that significant links overlap is a natural feature of coupled networks that coevolve together, and can emerge through different mechanisms [39, 120, 138].

Thus, it is important to develop mathematical tools beyond the mean filed approximation to study higher-order correlations which naturally emerge in coevolving coupled networks. While this has partially been done with the standard percolation process [63, 128], we should also consider other dynamics that better model the behaviour of coupled networks. For example, in such systems, important (high-degree) nodes are often better protected against failures than less important (low-degree) nodes, and thus it would interesting to consider a percolation process

where failure probability is inversely proportional to the degree of a node.

Finally, as we mentioned earlier, it might be the case that biological networks exhibit closer to random inter-network coupling than other networks, as a result of constraints and other limitations. This is definitely something that should be further explored. While for single networks, we know that certain constraints, such as space, highly affect the structure of networks (see for example Fig. 2.2), this issue has hardly been investigated for coupled networks, which are likely face other restrictions than single isolated networks.

CHAPTER SIX

EMPIRICAL STUDY OF COUPLED TRANSPORTATION NETWORKS

Robust and efficient transportation systems are critical to our society, playing an important role in human mobility, the exchange of goods and the spread of diseases. As such, and with the recent availability of massive data sets, the structure of transportation networks has been extensively studied using network science techniques, revealing some striking universal features, which have been the focus of many recent studies. However, an increasing number of cities are spanned by multiple interacting transportation networks, a fact that has been neglected by most studies so far. To fill the gap in the literature, in this chapter, we complement the theoretical work presented in previous chapters with a large-scale empirical study of interacting underground and street networks in the entire metropolitan areas of both London and New York. The goal of this chapter is therefore to explore the utility of coupled complex networks modelling, as well as demonstrating that they can deal with the scale and empirical complexity of real-world network exemplars. We find that coupling can strongly affect the structure and consequently the behaviour of multilayer transportation systems, suggesting that previous findings should be taken with care and calling for more empirical and theoretical studies on the topic. The chapter is based on the author's work [223].

6.1 Introduction

The networks discussed in previous chapters only existed in an abstract "network space" where the precise positions of the network nodes have no particular meaning. But many networks, including transportation and mobility networks, the Internet, mobile phone networks, power grids, social and contact networks, and neural networks, live in a real space. Such networks, called *spatial networks*, are networks for which the nodes are located in a space equipped with a metric, where for most practical applications, the space is the two-dimensional space and the metric is the usual Euclidean distance [26]. The links may either be spatially embedded, such as in roads or railway lines in transportation networks or cables in a power grid, or abstract entities, such as friendship relations in social networks, but in both cases, space intervenes in the fact that the connection probability between two nodes usually decreases with the distance between them, and this has important effects on the topological properties of spatial networks, and consequently on the processes which take place on them.

Spatial networks were actually long ago the subject of many studies in quantitative geography, aiming to understand the evolution of transportation networks and their role in sustainable urban forms [119]. But with the recent advances in our understanding of complex networks, and perhaps more importantly, the availability of datasets of large networks and larger computer capabilities, the interest in spatial networks was renewed in recent years, discovering surprising similarities in the connectivity patterns of a wide range of spatially embedded systems. For example, an immediate consequence of spatial constraints is the restriction of large degrees resulting in a shorter tail or even peaked degree distributions. Another consequence of the cost associated with the length of links is that long-range links must be justified for good economical reasons resulting in several non-linear correlations between topology, traffic and distance [26]. For example, empirical studies of the airline transportation network connecting airpots in the world through direct flights, have found super-linear scaling relations between airports' degrees, the length of links connected to them, and the amount of traffic going through those links, meaning that larger airports have longer connections, which also convey more traffic [16, 117, 118]. Similar behaviour was also found in the cargo ship network, where ports are connected through direct routes [129, 136]. Connection costs also favour the formation of cliques between spatially close nodes and thus increase the clustering coefficient [26]. Finally, the distribution of nodes is often not uniform in space, as in the case of the Internet, where the density of routers and autonomous systems has been shown to increase monotonically with the population density [246].

But unlike the networks mentioned above such as the Internet and airline networks, for many infrastructure networks, planarity is unavoidable. Roads, rail, and other transportation networks are spatial and to a good accuracy *planar networks*, meaning they can be drawn in the plane

in such a way that its links do not intersect (alternatively, they are forming nodes whenever two links cross) [26]. These networks possess many features similar to lattices: their degree distribution is extremely peaked and is thus of little interest, they are no longer small-worlds, i.e. the average shortest path between pairs of nodes scales faster than logarithmic with the system size, usually as $\sqrt{N}$, and the distribution of links length is rapidly decreasing [52, 59, 144, 152]. For many applications, planar spatial networks are the most important, and most studies have focused on these examples, and especially on street and road networks, where nodes represent street junctions and links represent the street segments connecting them [26]. These networks play a central role at the functionality and sustainability of cities, which are one of the most import self-organising complex systems to our society, especially as more than 50% of the world population lives today in cities and this figure is bound to increase [6].

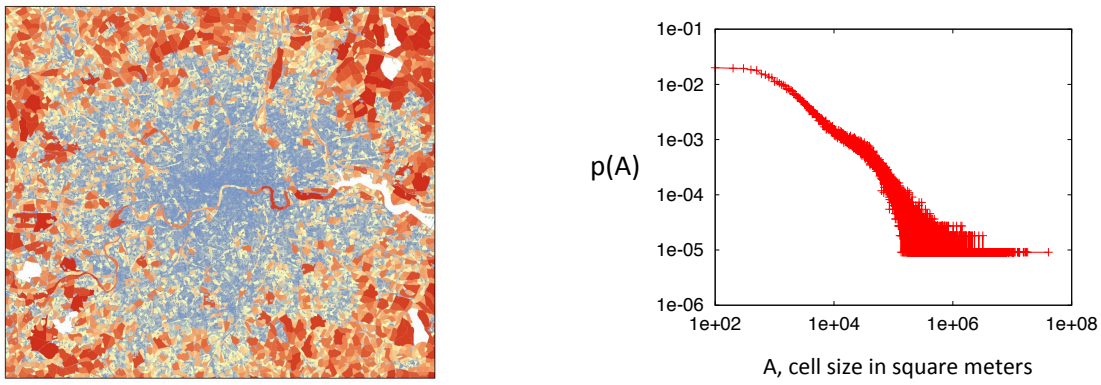


Figure 6.1: Spatial (left) and statistical (right) distribution of cells sizes, defined as the areas enclosed by roads, in London street networks. On the left we show a map of the metropolitan area in London where cells are coloured according to their size, and on the right we show the resulting distribution of cells sizes. In contrast to a regular lattice, the size of cells is broadly distributed, and often scales as $p(A) \sim A^{\alpha}$, with α close to 2 [143]. Note that the statistical distribution does not appear as a single line, as is often the case in real data, since there is a lot of variability at a small scale.

But although planarity strongly constrains their heterogeneity, urban street networks often display very distinctive features that are not captured in standard 2D models such as regular lattices. This could be a result of their self-organising structure – product of a stream of rational but usually uncoordinated decisions taking place through time [35, 196]. Remarkably, some irregular patterns emerge in the street patterns of very different cities that have been shaped by peculiar geographical, historical, and social-economical mechanisms. For example, the size of areas enclosed by the roads, called *cells*, have been found to display a broad distribution, in sharp contrast with the simple picture of an almost regular lattice, which would predict a distribution

peaked around the lattice edge length [143], see Fig. 6.1. The statistical and spatial distribution of betweenness centrality in street networks, characterising the importance of a node in the network (the more central the node and the larger the number of shortest paths going through this node), have also been found to differ from regular planar structures, with broad statistical distribution and distinctive spatial distribution where not only nodes close to the gravity centre have high centrality (as in lattice) [72, 73, 143], see Fig. 6.2. In particular, the betweenness centrality has been shown to highly correlate with microeconomic activities [189], urban growth [222] and land use intensity [236]. Other interesting features of street networks include their high local and global efficiency, compared to other planar networks with the same 'meshedness' (being equal to 0 for a tree and to 1 for a maximal planar graph) [52, 59]; varied accessibility accounting for the probability that an agent performing a self-avoiding random walk (i.e. does not visit the same point more than once) reaches a node after a certain number of steps [74]; and non-uniform distribution of nodes [28, 211].

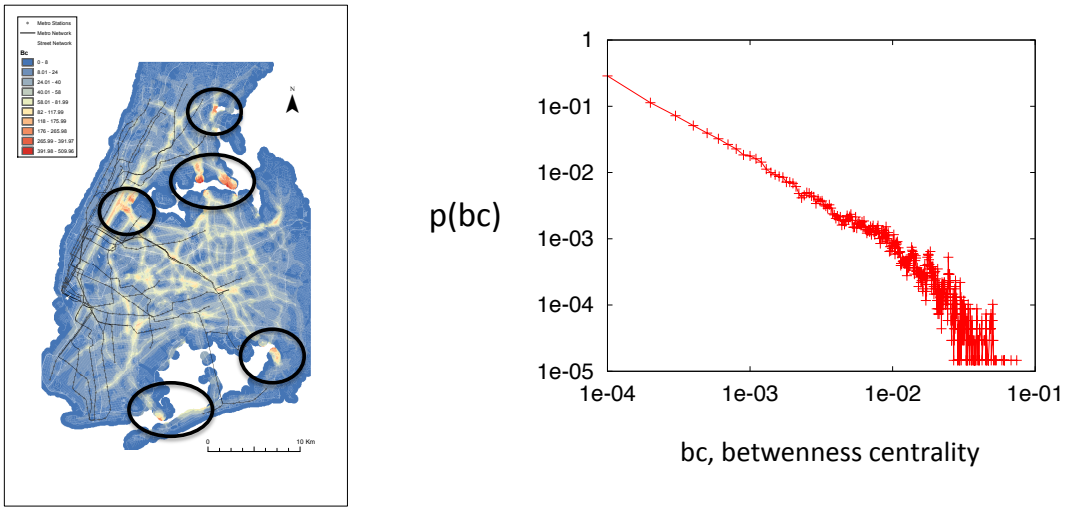


Figure 6.2: Spatial (left) and statistical (right) distribution of betweenness centrality in the street network of New York. Note that the high-centrality nodes are not necessarily concentrated around the geographical centre, as one would expect in a regular planar network. Specifically in New York, the bridges (circled in black) are the most central, but this structure may vary across different cities.

The common features found in empirical studies of street networks of different cities have attracted a lot of interest in recent years, and have partially been captured in various network formation models such as spatial Erdős-Rényi networks, and networks growth models with local optimisations rules. But despite these various advances, the study of spatial networks is not nearly as well-established as "standard" network science concepts and there are still many open problems, as discussed in the recent review by Barthélemy [26]. In particular, most large

cities today are spanned by more than one transportation systems, and underground systems are especially becoming popular, existing in all UN cities with above 10 million people [196]. The coexistence of multiple transportation systems in metropolitan areas can have a great impact on their behaviour, for example, by introducing congestions in streets close to underground stations and change the criticality of the system. However, this issue have hardly been dealt with in the complex networks literature, with no city-scale empirical studies yet published on the topic, as far as we aware, and very few theoretical ones [165, 166]. In the transportation science literature, traffic and congestion problems related to intermodality were largely considered, but these studies seems to ignore the relevance, in this context, of a correct topological analysis of the network's graph.

The aim of this chapter is therefore to fill the gap in the literature, providing a first empirical study of the structure of coupled transportation systems, while exploring the utility of coupled networks techniques to real-world problems. In particular, we analyse the interplay between the street and underground networks in the metropolitan areas of London and New York, examining how the existence of an underground system is affecting the structural properties, and consequently the behaviour of street networks. Both cities, although resulting from very different urban processes, are highly affected by the the nontrivial interactions between the transportation networks, which lead to great changing in their efficiency and robustness, both on local and global scales. The rest of this chapter is organised as follows: in section 6.2 we present the dataset from which we extract the coupled networks, compare their basic characteristics, and discuss some implementation details of the algorithm used to analyse the data. The main results of our empirical study are presented in section 6.3, while highlighting the commonality and differences between the two case studies of London and New York. We summarise out findings in section 6.4 and discuss possible directions in which they can be extended.

6.2 Data and methods

We exported the geo-referenced street and underground networks of New York and London from Open Street Map [2], see Fig 6.3(a)-(b), and following a series of data hygiene operations (e.g. combining multiple entries to underground stations, combining multiple railroad tracks), extracted two connected networks for each city (i.e. we only consider the largest connected component). Let $G_{\mathrm{s}} = (V_{\mathrm{s}}, E_{\mathrm{s}})$ denote a connected street network where street junctions (nodes) are connected by street segments (links), and $G_{\mathrm{u}} = (V_{\mathrm{u}}, E_{\mathrm{u}})$ denote a connected underground network where underground stations (nodes) are connected by railroad tracks (links). We define the coupled network $G_{\mathrm{c}} = (V_{\mathrm{s}} \cup V_{\mathrm{u}}, E_{\mathrm{s}} \cup E_{\mathrm{u}} \cup E_{\mathrm{c}})$, where E_{c} is the set of links connecting each underground station with its closest (in Euclidean distance) street junction, see illustration in

Fig 6.3(c).

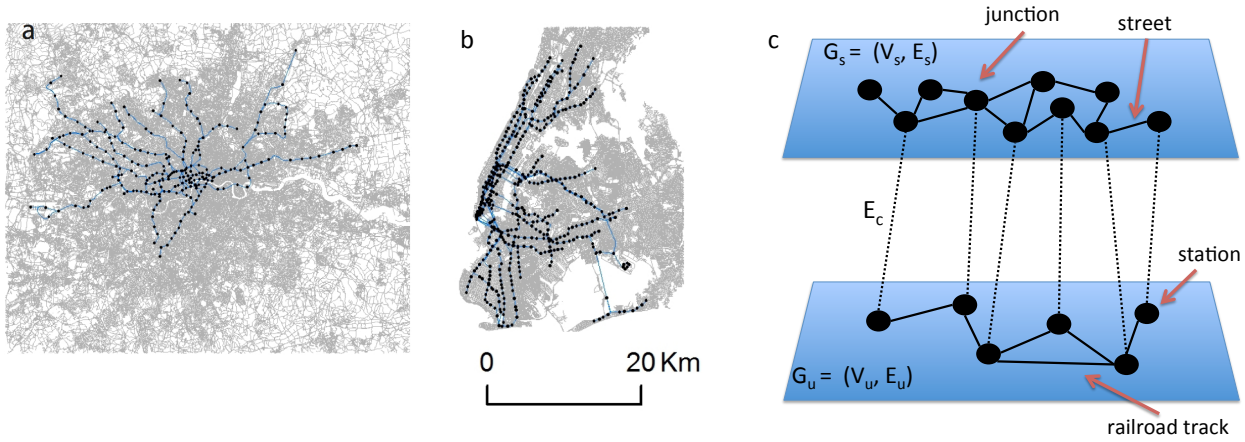


Figure 6.3: (a)-(b) The spatial extent of the two metropolitan cases (a) London and (b) New York. The two urban systems are intrinsically different where London is a typical mono-centric metropolitan region in which the underground developed branches in a semi-radial way from the city core to the peripheries, while New York presents an atypical shape with the street network mainly composed of separated islands connected by bridges and the underground. (c) schematic illustration of the resulting coupled network.

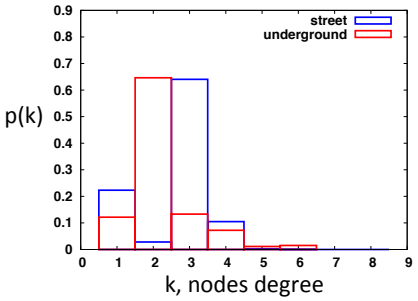
In Fig. 6.4, we show some basic characteristic of each network and the resulting degree distribution. First, in both cases of London and New York, the street network is significantly larger with 324536 nodes compared to 263 in London and 68417 compared to 454 in New York. Due to strong spatial constraints, the average degrees of the underground and street networks is relatively small, equal to 2.25 (underground) and 2.63 in London, and 2.15 (underground) and 3.29 in New York. The literature often divides cities into two categories, *self-organized cities*, such as Ahmedabad, Cairo, and Venice, grown throughout a largely self-organised, fine-grained historical process, out of the control of any central agency, and *planned cities*, such as Los Angeles, Richmond, and San Francisco, realized over a short period of time as the result of a single plan, usually exhibiting a regular grid like structure [59, 72]. London is a typical self-organised city, where most nodes have degree equal to 3, unlike New York where most nodes have degree equal to 4 (see degree distribution in Fig. 6.4), typical to grid-like structures. The underground as well is slightly different, where in New York, it is mostly organised as chains with more than 80% of underground stations have degree equal to 2. The cost, defined as the total length of links, of the underground system is only 1.1% percent in London and 3.4% in New York, but the average cost of links in the underground is much larger in both cases, where the average edge length is 0.85 km in New York and 1.3 km in London. This is due to the fact that the underground network is less dense, covering a large area with only a few hundreds of nodes

London

	N	M	Cost(km)	L (km)	Diam(km)
Street	324536	427920	34493.73	25.83	89.31
Underground	263	296	385.98	18.55	60.3
Coupled	324799	428479	34886.52	25.78	89.27

New York

	N	M	Cost(km)	L (km)	Diam(km)
Street	68417	112827	12153.81	17. 94	55.18
Underground	454	489	416.12	18.87	57.28
Coupled	68871	113770	12579.44	17.91	55.18



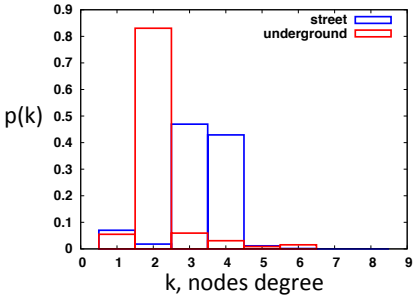


Figure 6.4: Basic quantitative analysis of London (top) New York (bottom) street and underground networks. The table shows the number of nodes, N, number of links, M, cost, defined as the total length of all links, average shortest path length, L and maximum shortest path length, *Diam*. The figures on the right show the resulting degree distribution.

and links. Finally, we show the average shortest path length, L and the maximum shortest path length, *Diam* between every pair of nodes. In spatial networks, the shortest path is a path with minimum length, in contrast to a path with minimum number of hops in topological (abstract) networks, and there is usually only one shortest path. Note that paths in the coupled network can traverse both underground and street links and are therefore always shorter than in the single networks. However, the gain of using underground links instead of only streets is not significant, 500 meters in London and 300 meters in New York on average, an expected result since the street network is very dense and thus follow closely underground paths.

Indeed, the efficiency of underground systems is rooted at their speed, compared to roads which are subject to congestions, and thus a distant place which is easily accessible through the underground appears to be closer than a near one with a longer time to access. Therefore, in the following, we measure the travel cost as time units, associating to underground links with a parameter $0 < \beta \leq 1$, meaning that the number of time units it takes to traverse an underground link of length l meters is βl, which is $\frac{1}{\beta}$ times faster than the time it takes to traverse the same length on the street network. Thus, smaller β corresponds to relatively faster underground speeds.

6.2.1 Computing quickest paths

In order to compute *quickest paths*, i.e. paths with a minimum number of time units, we associate each link $e = (i, j)$ in the coupled network with a weight w defined as

$$w(e) = \begin{cases} l(e), & \text{if } i, j \in V_{\text{s}} \\ \beta l(e), & \text{if } i, j \in V_{\text{u}} \\ d_e(i, j) & \text{otherwise.} \end{cases}$$

where $l(e)$ is the length of link e (i.e. length of a street segment or railroad track) and $d_e(i, j)$ is the Euclidean distance between i and j. Then, we run the Dijkstra's algorithm starting from each node in the network and obtain all the shortest paths between any pair of nodes. However, since the networks are quite large, one run of the Dijkstra's algorithm from a single source takes about 9.4 seconds for London and 1.6 for New York, meaning it would take more than a month (approximately 35 days) to compute all quickest paths in London for example. Therefore, we automatically divide the network into subsets of nodes, based on how many available cores are in the system at the moment, and each core is assigned with a set of nodes from which to start the network's crawl. For example, in a system of 64 cores, as we usually used here, a single run (i.e. for a given β value) takes around 13 hours for London and half an hour for New York. Note that each core still needs to store all the network in memory, but in our case it is only $300MB$ for New York and 1.28 GB for London, and thus does not constitute a problem. In larger networks, it might be possible to divide the network into "local areas" thanks to the fact that spatial planar networks are "large-worlds", i.e. there are no shortcuts which make it difficult to define the local environment of a node, but, once again, this was not an issue in our case. Finally, once all processes are finished, we collect the results and compute global quantities, as discussed in the next section.

6.3 Results

6.3.1 Quickest paths

Let $\tau_s(i, j)$ denote the travel cost (i.e. number of time units) of the quickest path between street nodes $i, j \in V_{\text{s}}$, and $\tau_c(i, j)$ correspond to the quickest path between i and j in the coupled network (i.e. a path which can traverse both street and underground links). The average travel cost to reach from a street node $i \in V_{\text{s}}$ to all other street nodes in the street and coupled network

respectively is denoted by

$$\tau_{\text{street}}(i) = \frac{1}{N_s - 1} \sum_{j \in V_s} \tau_s(i,j)$$
$$\tau_{\text{coupled}}(i) = \frac{1}{N_s - 1} \sum_{j \in V_s} \tau_c(i,j)$$

where $N_s = |V_s|$ is the number of street nodes. Note that we only count the cost of travel between i to other street nodes and not underground nodes, thus assuming that trip's targets are street junctions and not underground stations. This assumption will hold throughout this chapter, treating the underground as an auxiliary network, studying its effect on the coupled street network and not vice versa. Next, in order to quantify the time gained by using the underground, we define for each street node $i \in V_s$ the average ratio between the travel costs from i to other street networks through the coupled network and through the street network

$$q_{ms}(i) = \frac{1}{N_s - 1} \sum_{j \in V_s} \frac{\tau_c(i,j)}{\tau_s(i,j)} \tag{6.1}$$

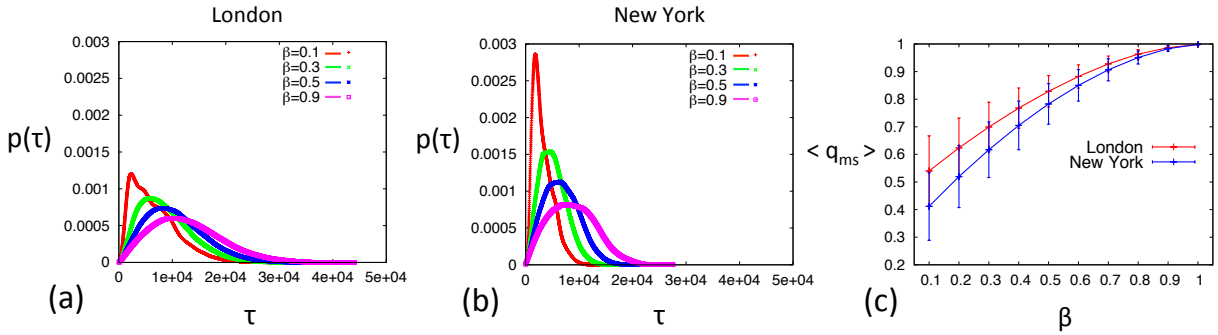


Figure 6.5: Distribution of quickest paths costs in the coupled network of London (a) and New York (b). (c) Average and standard deviation of q_{ms}, defined in Eq. 6.1, as a function of β.

In Fig. 6.5 we show the distribution of the cost of travel along quickest paths in the coupled network, τ_c, for different β, and the average q_{ms}, defined in Eq. 6.1, a function of β. First, we observe that as the underground becomes quicker (smaller β), the distribution of quickest paths is becoming more narrow and peaked meaning that more nodes are reachable within a small travel cost (higher peak) and all nodes are reachable within a medium travel cost (shorter tail). However, in New York, the change in the distribution is much more significant, with a 3.56 times higher peak for $\beta = 0.1$ than for $\beta = 1$, where in London the figure is only 2.033. This difference could be rooted at the fact that New York is composed of islands connected by underground links

and some congested bridges, and therefore when the underground is very efficient, then the city is "collapsing" into a very small place in terms of travel cost. Another reason is the extent of the areas considered, see Fig. 6.3, where New York is almost entirely covered by the underground network while in London the underground only spans the city centre. Finally, in Fig. 6.5(c) we show the average gain in terms of travel cost of having an underground system, as a function of its speed, β. When the underground network is only less than 1.5 times faster than the street network ($\beta = 0.7$), we already observe a gain of more than 10 percent. Considered that the cost of the underground network is about 1% percent of the cost of the street network (or 3% in New York), see Fig. 6.4, this is a very dramatic gain in terms of travel cost.

6.3.2 Interdependence

To quantify the added value of coupling to the reachability of nodes, in our case the contribution of the underground network to the reachability of street nodes, recent studies have introduced a quantity called *interdependence* [34, 165, 177]. The interdependence of a street node $i \in V_s$ is defined as

$$\lambda(i) = \frac{1}{N_s} \sum_{j \in V_s} \lambda(i,j) = \frac{1}{N_s} \sum_{j \in V_s} \frac{\sigma_{i,j}^{\text{coupled}}}{\sigma_{i,j}} \tag{6.2}$$

where $\sigma_{i,j}$ is the total number of quickest paths between i and j, among them $\sigma_{i,j}^{\text{coupled}}$ use links in both networks. Thus, interdependence is accounting for how much a street node is relying on the underground network to reach other street nodes. In Fig. 6.6 we show the average interdependence among all street nodes as a function of β (a) and the spatial distribution of

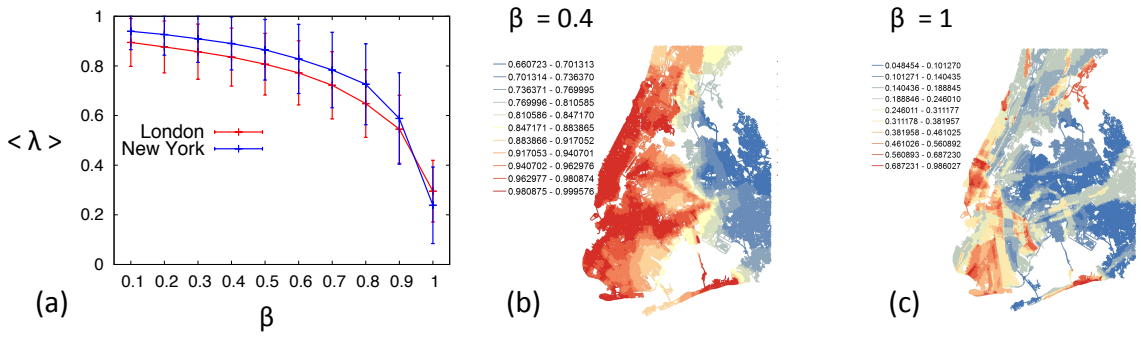


Figure 6.6: (a) Average and standard deviation of interdependence, as defined in 6.2, among all street nodes as a function of β. (b)-(c) Spatial distribution of interdependence in New York for $\beta = 0.4$ and $\beta = 1$ respectively.

interdependence in New York (b)-(c). In both cities, the existence of the underground has a very strong impact, for example, for $\beta = 0.8$, we obtain λ around 0.7, meaning that even when the underground is only 1.25 times faster than the street network, already about 70% percent of the quickest paths are going through the underground. We also observe a very quick increase for small speed ratio (larger β), with a more than 20% percent jump between $\beta = 1$ and $\beta = 0.9$, and a slightly smaller jump between from $\beta = 0.9$ to $\beta = 0.8$. Thus, in this range of β, a slight increase in the speed of the underground would already move 20% of the traffic to the underground, assuming that people travel along quickest paths, a point on which we further elaborate in the next section. Finally, from the spatial distribution of interdependence, see Fig. 6.6(b)-(c), we observe that even when there is no speed difference between the networks ($\beta = 1$), Manhattan for example, exhibits interdependence values larger than 0.5, meaning that more than 50% percent of the trips to Manhattan are *shorter* when using the underground. But we already know from previous results, see Fig. 6.5(c) for example, that on average these shortcuts are not significant. For $\beta = 0.4$, all places on the map have interdependence values larger than 0.66 meaning that even places not near the underground rely on its functionality in 66% of their quickest paths to other places.

In order to obtain a more detailed picture of nodes' interdependence at different scales, we define the interdependence profile

$$Q_\lambda(d) = \frac{1}{N(d)} \sum_{\substack{i,j \in V_s \\ d_e(i,j)=d}} \lambda(i,j) \tag{6.3}$$

where $d_e(i,j)$ is the Euclidean distance between i and j and $N(d)$ is the number of pairs of nodes

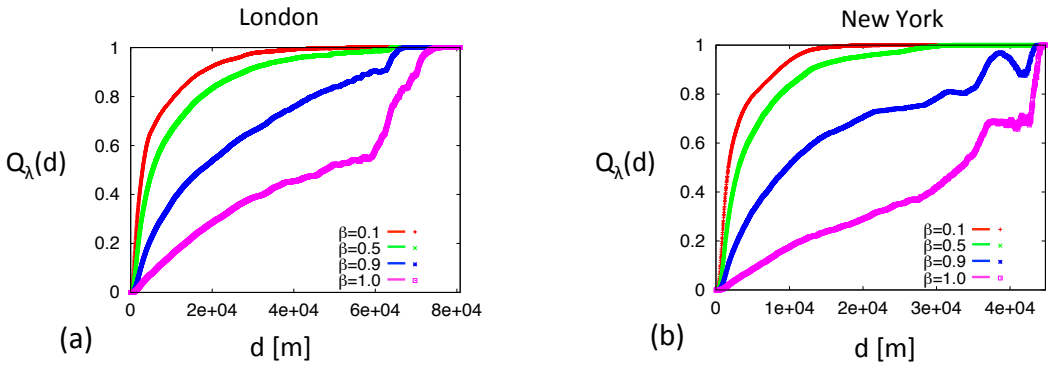


Figure 6.7: Interdependence profile, defined as the average interdependence of paths between nodes in a certain distance, see Eq. 6.3, in London (a) and New York (b).

at Euclidean distance d. In Fig. 6.6 we show the interdependence profile for various β in London (left) and New York (right). Both cities exhibit a very similar interdependence profile, especially for small β values, and given the different structure and evolution of these cities, this result might suggest a universal behaviour where the longer the trips (large d), the more beneficial it is to take the underground. As the rapidity factor gets larger, for example $\beta = 0.1$, the curve is going up very quickly meaning that already in very short distances, it is worth using the underground. In this case of $\beta = 0.1$, almost all trips longer than 1.5 kilometres in New York are faster when using the underground. If the underground is twice as fast ($\beta = 0.5$) than the street network, instead of 10 times faster ($\beta = 0.1$), then for trips longer than 2.9 kilometres it is almost always (99% percent) more cost effective to use the underground. The important point is that in both cities, medium range trips are almost always faster when using the underground, and when the rapidity of the underground is significantly larger compared to roads, e.g. twice as fast, then even for trips between near street junctions, "hopping" on the underground would make them faster.

6.3.3 Centralities

The results above suggest that the existence of an underground system has a very strong impact on the quickest paths in both cities, and therefore might lead changes in the distribution of traffic loads, and consequently to changes in the robustness and criticality of street networks. Betweenness centrality is one of the most well studied concept in networks in general and in spatial networks in particular, quantifying the importance of nodes in the networks as the amount of traffic going through them, assuming that traffic between all nodes is the same and only traversing shortest paths [72, 73, 97, 98]. Even if the underlying assumptions are not correct, the spatial distribution of the betweenness centrality gives important information about the coupling between space and the structure of the road network, and has been shown to highly correlate with micro economic activities [189], urban growth [222] and land use intensity [236]. We define the betweenness centrality of a street node $v \in V_s$ in the street network as

$$bc_s(v) = \frac{1}{\underbrace{(N_s - 1)(N_s - 2)}_{\text{number of pairs of street nodes}}} \sum_{i,j \in V_s} \frac{\sigma_{i,j}^{\text{street}}(v)}{\sigma_{i,j}^{\text{street}}} \tag{6.4}$$

where $\sigma_{i,j}^{\text{street}}$ is the total number of quickest paths between i and j in the street network, among them $\sigma_{i,j}^{\text{street}}(v)$ going through street node v. Similarly, we define the betweenness centrality of a

street node $v \in V_s$ in the coupled network as

$$bc_c(v) = \underbrace{\frac{1}{(N_s-1)(N_s-2)}}_{\text{number of pairs of street nodes}} \sum_{i,j \in V_s} \frac{\sigma_{i,j}^{\text{coupled}}(v)}{\sigma_{i,j}^{\text{coupled}}} \tag{6.5}$$

where $\sigma_{i,j}^{\text{coupled}}$ is the total number of quickest paths between i and j in the coupled network, among them $\sigma_{i,j}^{\text{coupled}}(v)$ going through street node v. In order to quantify the relative change in the betweenness centrality of street nodes with the introduction of an underground system, we define

$$\alpha(v) = \frac{bc_c(v) - bc_s(v)}{bc_s(v)}. \tag{6.6}$$

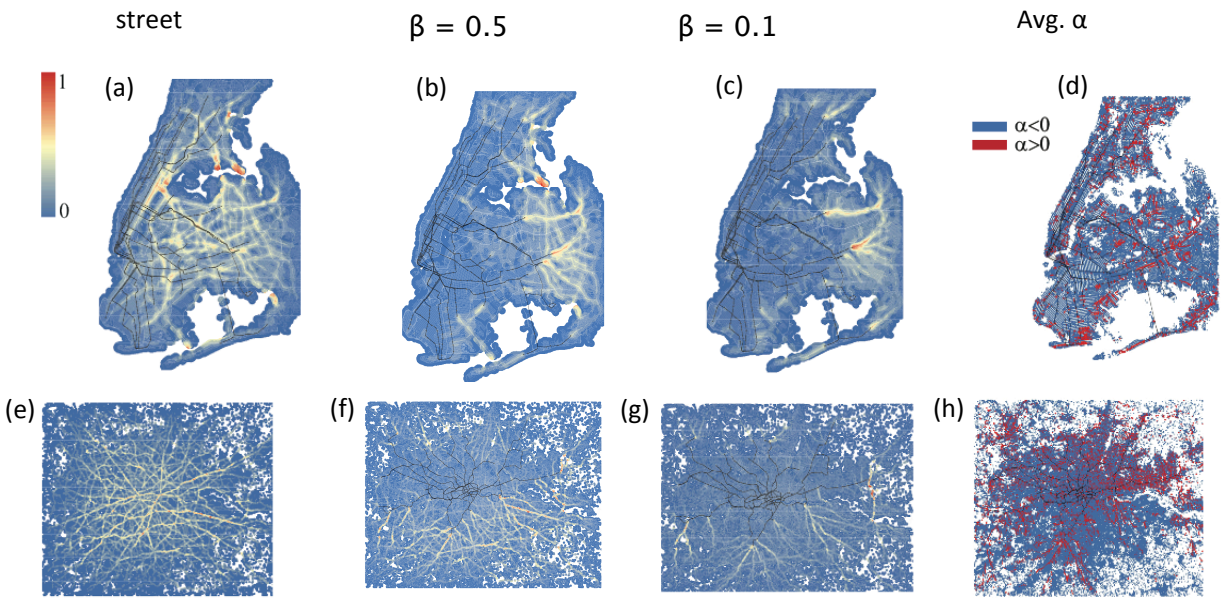


Figure 6.8: Spatial distribution of betweenness centrality of street nodes in New York (top) and London (bottom). (a),(e) betweenness centrality in the street network, bc_s. (b),(f) betweenness centrality in the coupled network, bc_c for $\beta = 0.5$. (c),(g) betweenness centrality in the coupled network, bc_c for $\beta = 0.1$. (d),(h) Average α, defined in 6.6, over all $\beta \in \{0.1, 0.2, \ldots, 1\}$.

In Fig. 6.8 we show the spatial distribution of betweenness centrality in the street network (a),(e), the coupled network for $\beta = 0.5$ (b),(f) and $\beta = 0.1$ (c),(g), and the relative increase in centrality α averaged over all $\beta \in \{0.1, 0.2, \ldots, 1\}$ (d),(h). These maps clearly display a dramatic change in the spatial distribution of central places when introducing an underground system, shifting

from internal street routes and bridges to intermodal places located at the terminal points of the underground networks, which are used as doors for suburban flows to reach the core urban areas. Remarkably, in both cities, these places are located in urban areas that do not overlap with the underground system, thus possibly creating congestions in totally unexpected places. In other words, the introduction of underground networks seem to create congestions in places located at the end of underground lines, and not for example, in the city centre, as one might expect. This can also been seen in the relative change in the centrality of nodes, see Fig. 6.8(d),(h), where blue areas are places in which the betweenness centrality decreases as a result of coupling, and red areas are the opposite. The maps show a complex process of loose-gain centrality in which the red nodes, that in average increase they centrality over all β, are spatially located in a very sparse manner and not only in the proximity of the underground network, suggesting that coupling can have a global effect on the distribution of traffic.

In order to quantify the effect of coupling on the criticality and robustness of street networks, we use to Gini coefficient to measure the statistical dispersion of the distribution of betweenness centrality in the coupled network as a function of β, see Fig. 6.9. The Gini coefficient $G \in [0,1]$,

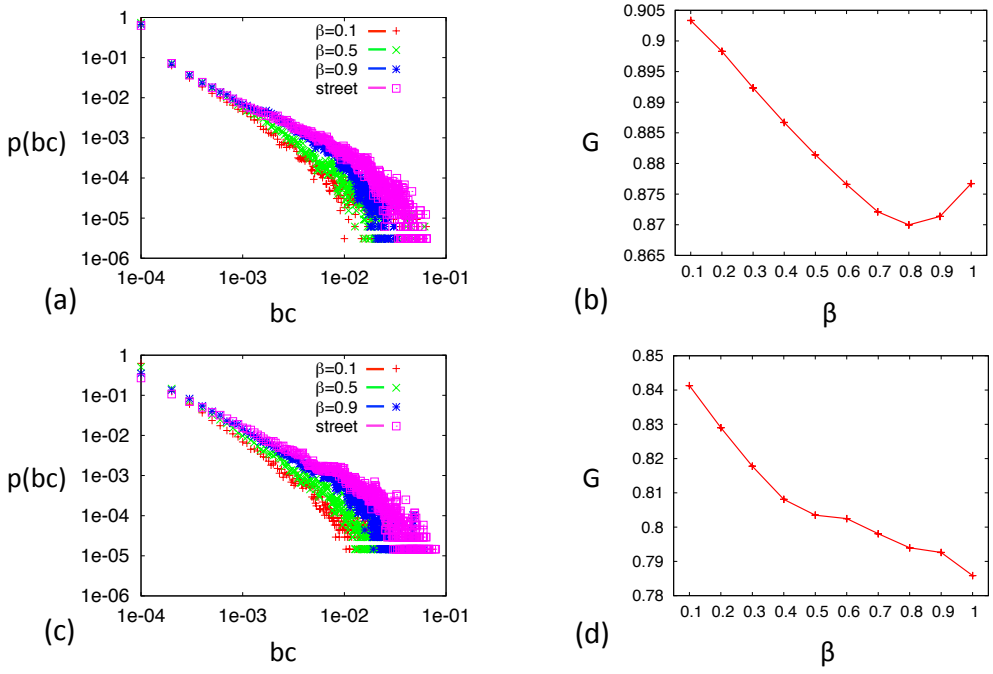


Figure 6.9: (a),(c) Distribution of betweenness centrality in the street network (pink) and in the coupled network for $\beta = 0.1$ (red), $\beta = 0.5$ (green) and $\beta = 0.9$ (blue) in London (top) and New York (bottom). (b),(d) Gini coefficient of the distribution of betweenness centrality in the coupled network, defined in 6.7, as a function of β.

typically used in economics for the purpose of describing the distribution of wealth within a nation, measures the inequality among values of a frequency distribution. In transportation systems, it is often used to characterise the disparity in the assignment of flows to the nodes or edges of a network, and is defined as [143, 165]

$$G = \frac{1}{2\,N_{\mathrm{s}}\,\langle bc_{\mathrm{c}}\rangle} \sum_{i,j\in V_{\mathrm{s}}} |bc_{\mathrm{c}}(i) - bc_{\mathrm{c}}(j)| \tag{6.7}$$

where $\langle bc_{\mathrm{c}}\rangle$ is the average betweenness centrality of street nodes in the coupled network. For example, if all flows were concentrated onto one node, G would be one, while if the flows were spread evenly across all nodes, G would be zero. In Fig. 6.9 we show the distribution of betweenness centrality and the resulting Gini coefficient in London (top) and New York (bottom). From both quantitates, we observe that as the rapidity of the underground increases (decreasing β), more quickest paths are moving to the underground, and the distribution is becoming less homogeneous, in agreement with the spatial distribution shown in Fig. 6.2, where we observed that intermodal points are becoming enormously more important. In other words, there is a trade-off between congestion on roads and congestion in intermodal points, and as cities increase their efficiency, at the same time, their robustness is decreasing since the traffic is concentrated at a very few intermodal places. In London, we observe a nontrivial optimal β for which flows are most homogeneously distributed across street junctions. In New York, however, there seem to be room for smaller and smaller β. Remarkably, this result was suggested in a recent theoretical model of coupled transportation networks, where depending on the distribution of trip's targets, two regimes were observed, one in which the optimal coupling is trivially the maximum, an another one where a nontrivial optimal coupling exists [165].

6.3.4 Local outreach

So far, we have considered the effect of coupling on global quantities such as quickest paths between *all* pair of nodes. In order to quantify the efficiency of coupled transportation networks on a local scale, we define the outreach of a street node $i \in V_{\mathrm{street}}$ as the average Euclidean distance from i to all other street nodes that are reachable within a given travel cost, τ^*

$$L(i) = \frac{1}{N(\tau^*)} \sum_{j|\tau_c(i,j)<\tau^*} d_e(i,j) \tag{6.8}$$

where $d_e(i,j)$ is the Euclidean distance between node i and j, and $N(\tau^*)$ is the number of nodes reachable within a given travel cost τ^*. Given a certain threshold τ^*, we define a subgraph containing all the street nodes, which their outreach is in the 4th quintile (i.e. approximately

20 percent of the nodes with the highest outreach), and count the number of component in the obtained subgraph, see Fig 6.10. As β decreases, the nodes having a high local outreach are concentrated at the centre, where the underground is most accessible, and the graph consisting of high-outreach nodes (red nodes on the map) is becoming less fragmented containing a smaller number of larger components. In other words, as the underground becoming more efficient, a continuous area of high-outreach nodes is emerging, meaning that a person can travel to this area (e.g. city centre) from far away places (large Euclidean distance) at a small travel cost, τ^*.

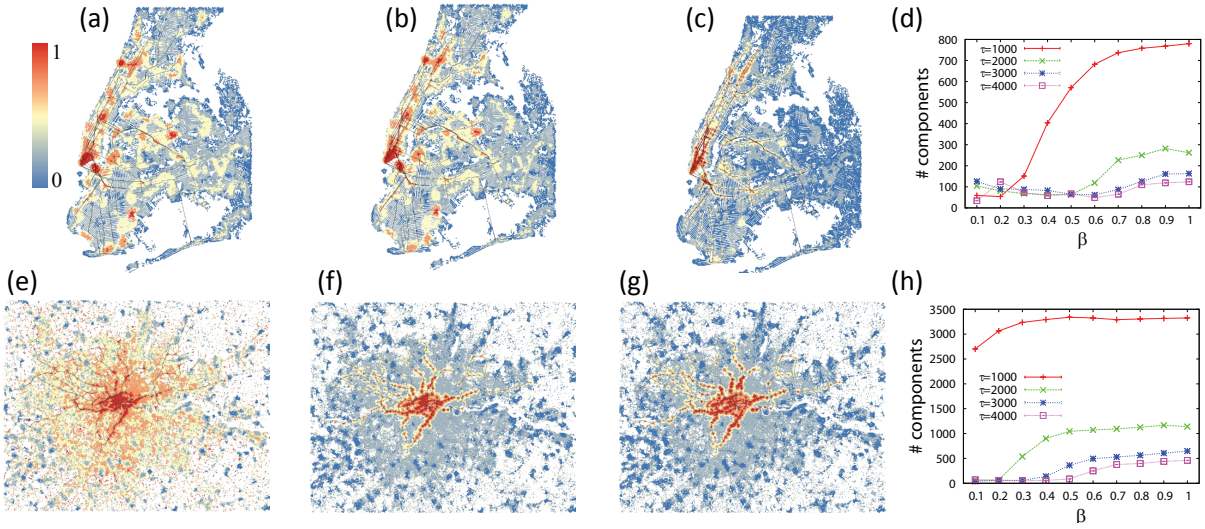


Figure 6.10: (a)-(c) Spatial distribution of local outreach in New York for travel cost $\tau^* = 1000$ and rapidity factor $\beta = 0.8$ (a), $\beta = 0.4$ (b) and $\beta = 0.2$ (c). (e)-(g) Spatial distribution of local outreach in London for travel cost $\tau^* = 2000$ and rapidity factor $\beta = 0.8$ (e), $\beta = 0.4$ (f) and $\beta = 0.2$ (g). (d),(h) Number of components in the graph containing only nodes with high outreach (red nodes in the maps) as a function of β in New York (top) and London (bottom).

Finally, in Fig 6.11, we show the average local outreach in London (left) and New York (right). As β increases the errors bars correspond to the standard deviation are becoming smaller, meaning that the distribution is becoming more peaked and almost all nodes can reach about the same Euclidean distance within a given travel cost. Also, larger travel costs result in a more dramatic decrease of the average with the increase of β, meaning that for large distances, the underground has a larger effect on the efficiency of trips, in agreement with the interdependence profile discussed in section 6.3.2.

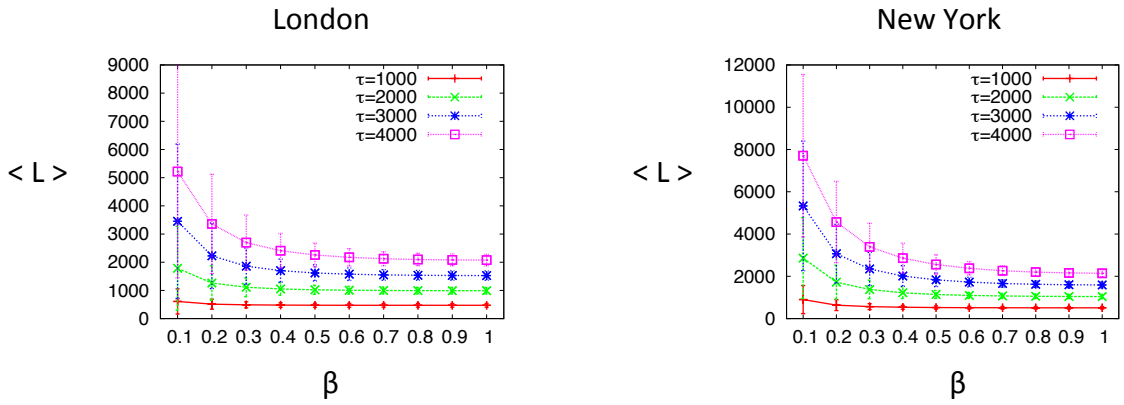


Figure 6.11: Average and standard deviation of local outreach in London (left) and New York (right).

6.4 Discussion

In this chapter, we have presented a large-scale empirical study of coupled transportation networks in the metropolitan areas of London and New York, two cities with very different characteristics resulting from different evolutionary processes. In particular, we have examined the effect of underground systems on the efficiency and robustness of street networks, which play a key role in the functionality and sustainability of cities and are thus of great importance to modern society. In order to account for the rapidity of underground networks compared to roads which are subject to congestions, we have associated underground links with a rapidity factor β meaning that the underground is $\frac{1}{\beta}$ times faster than the street network. In reality, the β value for New York can be approximated for example by 0.55 which is the average speed of underground trips, 17.4 mph [1], divided by the average speed of taxi trips over weekdays, 9.7 mph [3]. For London, the average speed of underground trips is 20.5 mph [4], compared with car trips, 10 mph [5], yielding $\beta = 0.48$, very similar o the one obtained for New York, and of the same order as the values considered in this work.

Our results suggest that coupling can have a great impact on the behaviour of street networks, and therefore must be taken into consideration in future studies, and perhaps even question the validity of well-established current studies. First, we find that the reachability of street junctions in both cities highly depends on the underground network, with close to 60% of quickest paths between street junctions are using the underground network already for $\beta = 0.9$, see Fig. 6.6. Moreover, the longer the trips, the more cost effective it is to "hop" on the underground, see Fig. 6.7, providing a mathematical support for our intuition.

Our most significant finding is the change in the distribution and dispersion of flows in street networks as a result of their coupling with faster underground systems. We have observed a trade-off between congestion on roads and congestion in intermodal points, which are the door for suburban flows to the city centre. As the rapidity of the underground is increased compared to roads, intermodal points are becoming enormously more important, subject to high congestions, and thus an attractive point for attacks on the system. In London, we have observed a nontrivial optimal speed ratio $\beta = 0.8$ (see Fig. 6.9), for which the flows are most evenly distributed across street junctions, and yet people can take advantage of a fast underground network, while in New York there seem to be a room for smaller and smaller β. Surprisingly, the existence of an underground network also seem to influence the betweenness centrality of places located in a very sparse manner and not only in the proximity of the underground network, see Fig. 6.8, suggesting a more complicated global effect which needs further investigation.

On a local scale, we discover that as β decreases, a continuous area of places that are reachable within a small travel cost (i.e. commutable zone) is emerging, see for example Manhattan in Fig. 6.10, creating opportunity for social and economic developments.

The work presented in this chapter provides the first empirical study on the effect of coupling on the structural properties of street networks. Using a coupled networks approach, we were able to account for the interplay between two very different transportation networks, street networks which are very dense and costly (in terms of total links length), and underground networks that are more sparse, efficient and usually concentrated mainly in the city centre. Studying these network separately, or alternatively treating them as one aggregated transportation network, would have provided a misleading topological picture, for example missing their most important nodes. By design, coupled networks contain more information than single isolated networks, and are bound to give a more detailed description of the systems they represent. This is obviously the cost that is associated with more comprehensive modelling techniques. In this chapter, we have shown the great return of using such approach, while demonstrating that it can also deal with the scale and empirical complexity of real-world network exemplars.

Our study raises many interesting research directions, that we hope would encourage more empirical and theoretical studies on the topic. First, to account for more realistic situations, it is important to consider constrained paths, for example, with only one subway line change, or only one change of transport mode. This can be accommodated within our framework by a version of $w(e)$ that is bounded in the number of changes of transport mode, or alternatively associating such routes with a relatively large cost.

Second, recent studies have consider simplest paths [235], those containing smallest number of

turns, instead of shortest paths or in our case quickest paths. This is also something that would be interesting to consider in the framework of coupled network, since underground networks often simplify the amount of information required to complete a trip. This is especially notable in the case of tourists which often use the underground not necessarily to save time, but just because it is simpler. As mentioned in sections 6.3.1-6.3.2, New York seems to be more affected by the existence of the underground, where more quickest paths are using the underground even on short trips. Since the underground is usually also simpler to use, this might explain the conception of people that New York is easier to get around than London.

Finally, perhaps the most important extension of our study would be to consider other traffic distributions, instead of the homogeneous assumption made when using betweenness centrality. This can be done either by using real traffic data, or by approximating the amount of traffic on an aggregated level (e.g. how many people are travelling from Manhattan to Queens every day) based on the population, and then apply the generalised betweenness centrality recently suggested by Carvalho *et al.* [60].

CHAPTER SEVEN

CONCLUSIONS

In this thesis, we have considered the structure and dynamics of coupled complex networks, an emerging and promising topic in network science aims at studying a wide range of complex multilayer heterogeneous systems. In the last few years, network scientists have been making a great collective effort to characterise and study the behaviour of interacting and interdependent networks, which are highly common in nature and engineering. These networks have often been found to behave very differently from isolated single networks, and thus may not be modelled as such. We have presented three different models, each tackling an important and timely challenge in the study of coupled networks, and lastly presented a large-scale study of real interacting transportation networks.

In chapter 3 we developed a theoretical framework to tackle a top network science challenge of understanding the robustness of modular or interconnected networks to attacks on interconnected and high betweenness centrality nodes. Our analytical formalism provides the critical concentration of interconnections between modules, above which the internal structure of each module is inseparable from the system as a whole. The applications of this formalism are wide spread and relevant to many real world complex systems in different disciplines. It can be also utilised for developing efficient immunisation strategies in geographically distant social networks, and evaluating the effectiveness of travel restrictions (which can be modelled as interconnections). Our formalism provides new insights into the resilience of social, biologic and economic systems, and provide critical information into how structure can affect the resilience of these systems, which in turn provides new strategies into planning and optimising such systems. Finally, the problem of attacks on high-betweenness nodes is very difficult to deal analytically and have only been studied numerically. Such attacks are considered to be among the most harmful attack strategies, thus providing new insights into the resilience and vulnerabilities of real world networks (i.e. infrastructure, transportation, internet, etc.).

In chapter 4 we considered the behaviour of coupled adaptive networks, describing a wide spectrum of systems, from brain networks consisting of coupled adaptive cortical areas, to social networks consisting of coupled adaptive distant populations. Using bifurcation analysis and extensive computer simulation, we find a new stable equilibrium that appears only in the case of weakly coupled networks and not at all in the case of a single adaptive network. Our analytical model provides the exact number of nodes that, if their state were synchronised across two networks (i.e. if they were coupled), a process spreading in one network would spread to the other. This approach could be used to determine optimum firewall placement to resist cyber attacks and to determine the maximum number of people that can travel between two countries such that an epidemic spreading in one will not spread to the other. Our work is also providing an important insight into the future efforts needed to take place in order to analytically describe the behaviour of coupled networks.

In chapter 5 we tackle a timely an ongoing question in the coupled networks literature, namely the effect of nonrandom correlated coupling on the behaviour of coupled systems. The vast majority of recent studies on this topic suggest that "standard" networks dynamics, such as percolation and epidemic spreading, are better operating on positively coupled networks than randomly or negatively coupled networks. While this result supports the strong positive correlations found in technological, social and economical networks, it fails to explain other network datasets, mainly ones extracted from biological data, which do not follow this seemingly optimal structure. Inspired by the robustness and stability usually found in biological systems, we suggest that the proposed optimal structure found in recent studies, is no longer optimal when considering that unlike connected nodes in a single network, coupled nodes play a "role" in more than one network, and therefore often share limited resources such as time, energy and memory. We studied a model of constraints epidemic spreading where nodes are limited to interact with a certain number of neighbours at each time step, and obtained qualitatively different results from previous studies, demonstrating the importance to carefully consider the new situations arise in coupled systems.

In chapter 6 we conducted a large-scale analysis of interacting spatial transportation networks in the entire metropolitan areas of both London and New York City. This is the first study of its kind analysing the interplay between the structure of mutually connected transportation networks, and their implications to multimodal cities, which are a critical element of social interaction. While filling an important gap in the literature, we explore the utility of coupled complex networks modelling, demonstrating that they can deal with the scale and empirical complexity of real-world network exemplars. We find that the nontrivial interaction patterns, arising in multilayer transportation systems, can have a profound effect on their behaviour,

demonstrating the insufficiency of existing models, which focus on single isolated transpiration networks: these are simply not sufficient to explain the complex urban forms observed today. Our study uncovers some of the most basic mechanisms affecting the stability, resilience and efficiency of cities, and could thus provide a great step towards design principles for optimal transportation networks.

7.0.1 Contributions

The contributions made in this thesis can be summarised as follows:

I A theoretical framework for studying the robustness of modular networks to attacks on interconnected and high betweenness centrality nodes. In particular, we provide a full solution for Erdős-Rényi modular networks.

II A theoretical framework to study epidemic spreading in coupled adaptive networks, using which we demonstrate that inter-network links between adaptive networks can give rise to nontrivial behaviours. In particular, we discover a new stationary state that only emerges in the case of weakly coupled networks.

III Numerical study of a constrained epidemic spreading model on correlated coupled networks, demonstrating that in contrast to recent studies on the topic, positively-correlated multiplexity structures are less spreading-efficient than negatively-correlated ones in the presence of constraints.

IV Large-scale empirical study of coupled underground and street networks in the metropolitan areas of London and New York, providing the first empirical result on the interplay between the topologies of coupled transportation networks, which significantly differs from previous results on isolated street networks.

7.0.2 Discussion

The results presented in this thesis illustrate the usability and application of coupled complex networks as a modelling framework for studying the behaviour of complex systems with separate but interacting concerns. In particular, the network perspective allows us to abstract complex dynamical and structural features into sets of microscopic local rules, which can then be studied numerically and analytically, to provide a macroscopic high-level description of the system. The more recent mathematical tools offered by coupled networks extend the standard models of isolated single networks to study richer and more realistic situations of interacting and

interdependent systems. However, it is often the question of how can one knows that the system in hand consists of one or multiple networks?

The standard, simple models often suffice to model system behaviour despite their strong assumptions. However many real-world systems have enough heterogeneous structure to disrupt these assumptions, making coupled networks an essential, complementary modelling framework. In chapter 3 for example, the multiple networks analytical approach allowed us to study the resilience of modular networks, consisting of tightly connected subnetworks (or modules) to the failure of the nodes connecting between the networks, which result in very different behaviour than the standard percolation model. In some cases, the study of interacting coexisting networks is crucial for understanding their impact on one another and thus their behaviour. In chapter 6 for example, we have shown that centrality of street junctions in large-scale street networks can not be explained without considering the interaction and dependency on other transportation systems, such as subway networks.

However, our analytical tools are still quite limited, and often fail to provide an accurate quantitative picture even in the most simple cases, such as homogeneous random networks, as we have seen in chapter 4: neither the mean field (zeroth order) approximation, nor the first order pair approximation, were enough to describe one of the observed equilibria in coupled adaptive networks. Dobson [84] compares such cases with the Halting Problem which essentially says that there are problems for which the code that computes the solution is the most concise description available. Numerical methods are indeed more flexible, constrained only by processing power and memory space, which can easily scale up to hundreds of thousands and even millions nodes using of–the–shelf libraries such NetwrokX and igraph. But although simple code can simulate the temporal and long-term behaviour of almost *any* process running on *any* network, it hardly provides an insight into why a given equilibrium behaviour occurs.

Thus, the pursuit for mathematical tools must continue, with the support of scaleable computational tools, which helps us to gain intuition and initial understanding about the questions we wish to answer. In a very short time, coupled networks have already been proved as a very useful abstraction of complex heterogeneous multilayer systems. Throughout this thesis, we have demonstrated the advantages of the coupled networks perspective to study modular, interacting, and multiplex systems, while tackling timely challenges in the complex networks literature. We strongly believe that the work presented in this thesis could open new avenues in understanding the interplay between the structure and function of complex coupled networks, which occur in all sects of today complex techno-social-economic world.

REFERENCES

[1] NYC transit forums, http://www.nyctransitforums.com/forums/topic/17313-subway-system-average-speed-by-line/.

[2] OpenStreetMap, http://www.openstreetmap.org/.

[3] The New York Times, http://www.nytimes.com/2010/03/24/nyregion/24traffic.html?_r=0.

[4] The Telegraph, http://www.telegraph.co.uk/travel/destinations/europe/uk/london/9789966/London-Underground-150-fascinating-Tube-facts.html.

[5] This Is Local London, http://www.thisislocallondon.co.uk/news/804876.london_cars_-move_no_faster_than_chickens/.

[6] UN population division., http://www.unpopulation.org.

[7] E. Ahmed, A. S. Elgazzar, and A. S. Hegazi. An overview of complex adaptive systems. *arXiv:nlin/0506059*, 2005.

[8] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, STOC '00, pages 171–180, New York, USA, 2000.

[9] M. Ajmone-Marsan, D. Arrowsmith, W. Breymann, O. Fritz, M. Masera, A. Mengolini, and A. Carbone. The emerging energy web. *Eur. Phys. J. Special Topics*, 214(1):547, 2012.

[10] R. Albert and A.-L. Barabási. Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.*, 85(24):5234, 2000.

[11] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47, 2002.

[12] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130, 1999.

[13] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378, 2000.

[14] A. F. Alexander-Bloch, N. Gogtay, D. Meunier, R. Birn, L. Clasen, F. Lalonde, R. Lenroot, J. Giedd, and E. T. Bullmore. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in systems neuroscience*, 4:147, 2010.

[15] A. Allard, P.-A. Noël, L. J. Dubé, and B. Pourbohloul. Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Phys. Rev. E*, 79(3):036113, 2009.

[16] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. U.S.A.*, 97(21):11149, 2000.

[17] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1992.

[18] J. P. Bagrow, S. Lehmann, and Y.-Y. Ahn. Robustness and modular structure in networks. *arXiv:1102.5085*, 2011.

[19] N. Bailey. *The Mathematical Theory of Infectious Diseases and Its Applications*. 1975, New York, Hafner Press.

[20] A. L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311(3-4):590, 2002.

[21] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[22] A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281(1-4):69, 2000.

[23] A.-L. Barabási and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, 5(2):101, 2004.

[24] L. Bargigli, G. di Iasio, L. Infante, F. Lillo, and F. Pierobon. The multiplex structure of interbank networks. *arXiv:1311.4798*, 2013.

[25] M. Barigozzi, G. Fagiolo, and D. Garlaschelli. Multinetwork of international trade: A commodity-specific analysis. *Phys. Rev. E*, 81(4):046104, 2010.

[26] M. Barthélemy. Spatial networks. *Physics Reports*, 499(1–3):1, 2011.

[27] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *J. Theor. Biol.*, 235(2):275, 2005.

[28] M. Barthélemy and A. Flammini. Co-evolution of density and topology in a simple model of city formation. *Netw. Spatial Econ.*, 9(3):401, 2009.

[29] P. Barttfeld, B. Wicker, S. Cukier, S. Navarta, S. Lew, R. Leiguarda, and M. Sigman. State-dependent changes of connectivity patterns and functional brain network topology in autism spectrum disorder. *Neuropsychologia*, 50(14):3653, 2012.

[30] A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin. The extreme vulnerability of interdependent spatially embedded networks. *Nat. Phys.*, 9(10):667, 2013.

[31] D. S. Bassett and M. S. Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200, 2011.

[32] M. Bastian, S. Heymann, and M. Jacomy. Gephi: an open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*, 2009.

[33] V. Batagelj and U. Brandes. Efficient generation of large random networks. *Phys. Rev. E*, 71(3):036113, 2005.

[34] F. Battiston, V. Nicosia, and V. Latora. Structural measures for multiplex networks. *Phys. Rev. E*, 89(3):032804, 2014.

[35] M. Batty. *Cities and Complexity*. MIT Press, 2007.

[36] E. A. Bender and E. Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296, 1978.

[37] Y. Berezin, A. Bashan, M. M. Danziger, D. Li, and S. Havlin. Spatially localized attacks on interdependent networks: the existence of a finite critical attack size. *arXiv:1310.0996*, 2013.

[38] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U.S.A.*, 104(17):7301, 2007.

[39] G. Bianconi. Statistical mechanics of multiplex networks: Entropy and overlap. *Phys. Rev. E*, 87(5):062806, 2013.

[40] G. Bianconi and S. N. Dorogovtsev. Multiple percolation transitions in a configuration model of network of networks. *arXiv:1402.0218*, 2014.

[41] G. Bianconi, S. N. Dorogovtsev, and J. F. F. Mendes. Mutually connected component of network of networks. *arXiv:1402.0215*, 2014.

[42] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4–5):175, 2006.

[43] G. A. Böhme and T. Gross. Analytical calculation of fragmentation transitions in adaptive networks. *Phys. Rev. E*, 83(3):035101, 2011.

[44] B. Bollobás. *Modern Graph Theory*. Graduate texts in mathematics. Springer, 1998.

[45] B. Bollobás. *Random Graphs*. Cambridge University Press, 2nd edition, 2001.

[46] B. Bollobás, R. Kozma, and D. Miklos. *Handbook of Large-Scale Random Networks: Bolyai Society Mathematical Studies*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[47] S. A. Boorman and H. C. White. Social structure from multiple networks. II. Role structures. *Am. J. Sociol.*, 81(6):1384, 1976.

[48] S. Bornholdt and T. Rohlf. Topological evolution of dynamical networks: Global criticality from local dynamics. *Phys. Rev. E*, 84(26):6114, 2000.

[49] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Comput. Netw.*, 33(1-6):309, 2000.

[50] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht. Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci. U.S.A.*, 109(12):E680, 2012.

[51] C. D. Brummitt, K.-M. Lee, and K.-I. Goh. Multiplexity-facilitated cascades in networks. *Phys. Rev. E*, 85(4):045102, 2012.

[52] J. Buhl, J. Gautrais, N. Reeves, R. V. Solé, S. Valverde, P. Kuntz, and G. Theraulaz. Topological patterns in street networks of topological patterns in street networks of self-organized urban settlements. *Eur. Phys. J. B*, 49(4):513, 2006.

[53] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025, 2010.

[54] S. V. Buldyrev, N. W. Shere, and G. A. Cwilich. Interdependent networks with identical degrees of mutually dependent nodes. *Phys. Rev. E*, 83(1):016112, 2011.

[55] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186, 2009.

[56] E. Bullmore and O. Sporns. The economy of brain network organization. *Nat. Rev. Neurosci.*, 13(5):336, 2012.

[57] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85(25):5468, 2000.

[58] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. d. Pozo, and S. Boccaletti. Emergence of network features from multiplexity. *Sci. Rep.*, 3(1344), 2013.

[59] A. Cardillo, S. Scellato, V. Latora, and S. Porta. Structural properties of planar graphs of urban street patterns. *Phys. Rev. E*, 73(6):066107, 2006.

[60] R. Carvalho, L. Buzna, F. Bono, E. Gutiérrez, W. Just, and D. Arrowsmith. Robustness of trans-European gas networks. *Phys. Rev. E*, 80(1):016106, 2009.

[61] R. Carvalho, L. Buzna, F. Bono, M. Masera, D. K. Arrowsmith, and D. Helbing. Resilience of natural gas networks during conflicts, crises and disruptions. *PLoS ONE*, 9(3):e90265, 2014.

[62] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81(2):591, 2009.

[63] D. Cellai, E. López, J. Zhou, J. P. Gleeson, and G. Bianconi. Percolation in multiplex networks with overlap. *Phys. Rev. E*, 88(5):052811, 2013.

[64] W.-K. Cho, K.-I. Goh, and I.-M. Kim. Correlated couplings and robustness of coupled networks. *arXiv:1010.4971*, 2010.

[65] J. Clarke. Contact tracing for chlamydia: data on effectiveness. *International Journal of STD & AIDS*, 9(4):187, 1998.

[66] A. Clauset, C. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661, 2009.

[67] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.*, 85(21):4626, 2000.

[68] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Phys. Rev. Lett.*, 90(5):058701, 2003.

[69] R. Cohen, S. Havlin, and D. ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91(24):247901, 2003.

[70] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. U.S.A.*, 103(7):2015, 2006.

[71] E. Cozzo, A. Arenas, and Y. Moreno. Stability of boolean multilevel networks. *Phys. Rev. E*, 86(3):036115, 2012.

[72] P. Crucitti, V. Latora, and S. Porta. Centrality in networks of urban streets. *Chaos*, 16(1):015113, 2006.

[73] P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *Phys. Rev. E*, 73(3):036125, 2006.

[74] L. da F. Costa, B. A. N. Travençolo, M. P. Viana, and E. Strano. On the efficiency of transportation systems in large cities. *Europhys. Lett.*, 91(1):18003, 2010.

[75] M. Danziger, A. Bashan, Y. Berezin, and S. Havlin. Interdependent spatially embedded networks: Dynamics at percolation threshold. In *Conference on Signal-Image Technology Internet-Based Systems (SITIS), 2013 International*, pages 619–625, 2013.

[76] D. Davis, R. Lichtenwalter, and N. Chawla. Multi-relational link prediction in heterogeneous information networks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2011*, pages 281–288, 2011.

[77] F. C. Davis, A. R. Knodt, O. Sporns, B. B. Lahey, D. H. Zald, B. D. Brigidi, and A. R. Hariri. Impulsivity and the modular organization of resting-state neural networks. *Cerebral Cortex*, 23(6):1444, 2013.

[78] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas. Mathematical formulation of multilayer networks. *Phys. Rev. X*, 3(4):041022, 2013.

[79] D. J. de Solla Price. Networks of scientific papers. *Science*, 149(3683):510, 1965.

[80] G. Demirel, F. Vazquez, G. Böhme, and T. Gross. Moment-closure approximations for discrete adaptive networks. *Physica D*, 267(0):68, 2014.

[81] Z. Dezső and A.-L. Barabási. Halting viruses in scale-free networks. *Phys. Rev. E*, 65(5):055103, 2002.

[82] M. Dickison, S. Havlin, and H. E. Stanley. Epidemics on interconnected networks. *Phys. Rev. E*, 85(6):066109, 2012.

[83] A.-L. Do, L. Rudolf, and T. Gross. Patterns of cooperation: fairness and coordination in networks of interacting agents. *New J. Phys.*, 12:063023, 2010.

[84] S. Dobson. *The edge of computer science*. http://www.simondobson.org/2013/07/edge/, 2014.

[85] J. Donges, H. Schultz, N. Marwan, Y. Zou, and J. Kurths. Investigating the topology of interacting networks. *Eur. Phys. J. B*, 84(4):635, 2011.

[86] S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, and A. Y. Zyuzin. Organization of modular networks. *Phys. Rev. E*, 78(5):056106, 2008.

[87] H. Ebel and S. Bornholdt. Coevolutionary games on networks. *Phys. Rev. E*, 66(5):056118, 2002.

[88] V. M. Eguíluz and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.*, 89(10):108701, 2002.

[89] P. Erdős and A. Rényi. On random graphs. *Publ. Math. Debrecen*, 6:290, 1959.

[90] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17, 1960.

[91] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the Internet. *Phys. Rev. Lett.*, 90(14):148701, 2003.

[92] B. Ermentrout. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. SIAM, Philadelphia, 2002.

[93] L. Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128, 1736.

[94] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. *SIGCOMM Comput. Commun. Rev.*, 29:251, 1999.

[95] M. R. Fitzgerald, D. Thirlby, and C. A. Bedford. The outcome of contact tracing for gonorrhoea in the United Kingdom. *International Journal of STD & AIDS*, 9(11):657, 1998.

[96] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75, 2010.

[97] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35, 1977.

[98] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215, 1979.

[99] S. Funk, M. Salathé, and V. A. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface*, 7(50):1247, 2010.

[100] A. Galstyan and P. Cohen. Cascading dynamics in modular networks. *Phys. Rev. E*, 75(3):036109, 2007.

[101] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Robustness of a network formed by interdependent networks with a one-to-one correspondence of dependent nodes. *Phys. Rev. E*, 85(6):066134, 2012.

[102] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Robustness of a network of networks. *Phys. Rev. Lett.*, 107(19):195701, 2011.

[103] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin. Networks formed from interdependent networks. *Nat. Phys.*, 8(1):40, 2012.

[104] A. Garas, P. Argyrakis, and S. Havlin. The structural role of weak and strong links in a financial market network. *Eur. Phys. J. B*, 63(2):265, 2008.

[105] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340, 1977.

[106] J. P. Gleeson. Cascades on correlated and modular random networks. *Phys. Rev. E*, 77(046117), 2008.

[107] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.*, 110(2):028701, 2013.

[108] J. Gómez-Gardeñes, C. Gracia-Lázaro, L. M. Floría, and Y. Moreno. Evolutionary dynamics on interdependent populations. *Phys. Rev. E*, 86(5):056113, 2012.

[109] J. Gómez-Gardeñes, I. Reinares, A. Arenas, and L. M. Floria. Evolution of cooperation in multiplex networks. *Sci. Rep.*, 2(620), 2012.

[110] C. Granell, S. Gómez, and A. Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys. Rev. Lett.*, 111(12):128701, 2013.

[111] P. Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157, 1983.

[112] T. Gross, C. D'Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.*, 96(20):208701, 2006.

[113] T. Gross and I. G. Kevrekidis. Coarse-graining adaptive coevolutionary network dynamics via automated moment closure. *Europhys. Lett.*, 82:380041, 2008.

[114] T. Gross. Interplay of network state and topology in epidemic dynamics. In S. Boccaletti, V. Latora, and Y. Moreno, editors, *Handbook on Biological Networks*, volume 10 of *World Scientific Lecture Notes in Complex Systems*. World Scientific, 2009.

[115] T. Gross and B. Blasius. Adaptive coevolutionary networks: a review. *J. R. Soc. Interface*, 5(10):259, 2008.

[116] T. Gross and H. Sayama. *Adaptive Networks: Theory, Models and Applications*. Understanding Complex Systems. Springer Berlin Heidelberg, 2009.

[117] R. Guimerá and L. Amaral. Modeling the world-wide airport network. *Eur. Phys. J. B*, 38(2):381, 2004.

[118] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. U.S.A.*, 102(22):7794, 2005.

[119] P. Haggett and R. J. Chorley. *Network Analysis in Geography*. Edward Arnold, London, 1969.

[120] A. Halu, S. Mukherjee, and G. Bianconi. Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles. *Phys. Rev. E*, 89(1):012806, 2014.

[121] J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430(6995):88, 2004.

[122] S. Havlin, D. Kenett, E. Ben-Jacob, A. Bunde, R. Cohen, H. Hermann, J. Kantelhardt, J. Kertész, S. Kirkpatrick, J. Kurths, J. Portugali, and S. Solomon. Challenges in network science: Applications to infrastructures, climate, social systems and economics. *Eur. Phys. J. Special Topics*, 214(1):273, 2012.

[123] Y. He, J. Wang, L. Wang, Z. J. Chen, C. Yan, H. Yang, H. Tang, C. Zhu, Q. Gong, Y. Zang, and A. C. Evans. Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS ONE*, 4(4):e5226, 2009.

[124] H. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599, 2000.

[125] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65(5):056109, 2002.

[126] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3):97, 2012.

[127] Y. Hu, B. Ksherim, R. Cohen, and S. Havlin. Percolation in interdependent and interconnected networks: Abrupt change from second- to first-order transitions. *Phys. Rev. E*, 84(6):066116, 2011.

[128] Y. Hu, D. Zhou, R. Zhang, Z. Han, C. Rozenblat, and S. Havlin. Percolation of interdependent networks with intersimilarity. *Phys. Rev. E*, 88(5):052805, 2013.

[129] Y. Hu and D. Zhu. Empirical analysis of the worldwide maritime transportation network. *Physica A*, 388(10):2061, 2009.

[130] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Robustness of interdependent networks under targeted attack. *Phys. Rev. E*, 83(6):065101, 2011.

[131] X. Huang, S. Shao, H. Wang, S. V. Buldyrev, H. E. Stanley, and S. Havlin. The robustness of interdependent clustered networks. *Europhys. Lett.*, 101(1):18002, 2013.

[132] F. hung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. U.S.A.*, 99(25):15879, 2002.

[133] J. Ito and K. Kaneko. Spontaneous structure formation in a network of dynamic elements. *Phys. Rev. E*, 67(4):046226, 2003.

[134] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41, 2001.

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

[136] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *J. R. Soc. Interface*, 7(48):1093, 2010.

[137] M. J. Keeling and K. T. Eames. Networks and epidemic models. *J. R. Soc. Interface*, 2(4):295, 2005.

[138] J. Y. Kim and K.-I. Goh. Coevolution and correlated multiplexity in multiplex networks. *Phys. Rev. Lett.*, 111(5):058702, 2013.

[139] M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *arXiv:1309.7233*, 2013.

[140] Y. Kornbluth, S. Lowinger, G. Cwilich, and S. V. Buldyrev. Cascading failures in networks with proximate dependent nodes. *Phys. Rev. E*, 89(3):032808, 2014.

[141] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85(21):4629, 2000.

[142] Y. A. Kuznetsov. *Elements of Applied Bifurcation Theory*, volume 112 of *Applied Mathematical Sciences*. Springer-Verlag, 3rd edition, 1995.

[143] S. Lämmer, B. Gehlsen, and D. Helbing. Scaling laws in the spatial structure of urban road networks. *Physica A*, 363(1):89, 2006.

[144] V. Latora and M. Marchiori. Is the Boston subway a small-world network? *Physica A*, 314(1–4):109, 2002.

[145] K.-M. Lee, J. Y. Kim, W. kuk Cho, K.-I. Goh, and I.-M. Kim. Correlated multiplexity and connectivity of multiplex random networks. *New J. Phys.*, 14(3):033027, 2012.

[146] E. A. Leicht and R. M. D'Souza. Percolation on interacting networks. *arXiv:0907.0894*, 2009.

[147] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 641–650, New York, 2012. ACM.

[148] W. Li, A. Bashan, S. V. Buldyrev, H. E. Stanley, and S. Havlin. Cascading failures in interdependent lattice networks: The critical role of the length of dependency links. *Phys. Rev. Lett.*, 108(22):228702, 2012.

[149] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, and M. Murray. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966, 2003.

[150] A. L. Lloyd and R. M. May. How viruses spread among computers and people. *Science*, 292(5520):1316, 2001.

[151] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé. Adaptive networks: Coevolution of disease and topology. *Phys. Rev. E*, 82(3):036116, 2010.

[152] A. P. Masuccia, D. Smith, A. Crooks, and M. Batty. Random planar graphs and the London street network. *Eur. Phys. J. B*, 71:259-, 2009.

[153] R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E*, 64(6):066112, 2001.

[154] C. Meisel, A. Storch, S. Hallmeyer-Elgner, E. Bullmore, and T. Gross. Failure of adaptive self-organized criticality during epileptic seizure attacks. *PLoS Comput. Biol.*, 8(1):e1002312, 2012.

[155] S. Melnik, M. A. Porter, P. J. Mucha, and J. P. Gleeson. Dynamics on modular networks with heterogeneous correlations. *Chaos*, 24:023106, 2014.

[156] D. Meunier, S. Achard, A. Morcom, and E. Bullmore. Age-related changes in modular organization of human brain functional networks. *NeuroImage*, 44(3):715, 2009.

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

[158] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv:cond-mat/0312028*, 2003.

[159] B. Min, S. D. Yi, K.-M. Lee, and K.-I. Goh. Network robustness of correlated multiplex networks. *arXiv:1307.1253*, 2013.

[160] M. Mitchell. *Complexity: A Guided Tour*. Oxford University Press, USA, 2011.

[161] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161, 1995.

[162] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Comb. Probab. Comput.*, 7(3):295, 1998.

[163] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B*, 26(4):521, 2002.

[164] Y. Moreno, M. Nekovee, and A. F. Pacheco. Dynamics of rumor spreading in complex networks. *Phys. Rev. E*, 69:066130, 2004.

[165] R. G. Morris and M. Barthélemy. Transport on coupled spatial networks. *Phys. Rev. Lett.*, 109(12):128703, 2012.

[166] R. Morris and M. Barthélemy. Spatial effects: Transport on interdependent networks. In G. D'Agostino and A. Scala, editors, *Networks of Networks: The Last Frontier of Complexity*, Understanding Complex Systems, pages 145–161. Springer International Publishing, 2014.

[167] T. Nakata. Collision probability for an occupancy problem. *Statistics & Probability Letters*, 78(13):1929, 2008.

[168] C. L. Nehaniv. Asynchronous automata networks can emulate any synchronous automata network. *International Journal of Algebra and Computation*, 14:719, 2004.

[169] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167, 2003.

[170] M. E. J. Newman. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E*, 64(1):016131, 2001.

[171] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.

[172] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, 67(2):026126, 2003.

[173] M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.*, 103(23):8577, 2006.

[174] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113, 2004.

[175] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64(2):026118, 2001.

[176] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, 2010.

[177] V. Nicosia, G. Bianconi, V. Latora, and M.Barthélemy. Growing multiplex networks. *Phys. Rev. Lett.*, 111(5):058701, 2013.

[178] A. Noack. Modularity clustering is force-directed layout. *Phys. Rev. E*, 79(2):026102, 2009.

[179] R. Parshani, C. Rozenblat, D. Ietri, C. Ducruet, and S. Havlin. Inter-similarity between coupled networks. *Europhys. Lett.*, 92(6):68002, 2010.

[180] 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. *Phys. Rev. Lett.*, 105(4):048701, 2010.

[181] R. Parshani, S. V. Buldyrev, and S. Havlin. Critical effect of dependency groups on the function of networks. *Proc. Natl. Acad. Sci. U.S.A.*, 108(3):1007, 2011.

[182] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.*, 87(25):258701, 2001.

[183] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, 63(6):066117, 2001.

[184] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200, 2001.

[185] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E*, 65(3):035108, 2002.

[186] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65(3):036104, 2002.

[187] N. Peyrard, U. Dieckmann, and A. Franc. Long-range correlations improve understanding of the influence of network structure on contact dynamics. *Theoretical Population Biology*, 73(3):383, 2008.

[188] M. J. O. Pocock, D. M. Evans, and J. Memmott. The robustness and restoration of a network of ecological networks. *Science*, 335(6071):973, 2012.

[189] S. Porta, V. Latora, F. Wang, E. Strano, A. Cardillo, S. Scellato, V. Iacoviello, and R. Messora. Street centrality and densities of retail and services in Bologna, Italy. *Env. Plan. B*, 36(3):450, 2009.

[190] D. D. S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27(5):292, 1976.

[191] F. Radicchi. Driving interconnected networks to supercriticality. *Phys. Rev. X*, 4(2):021014, 2014.

[192] F. Radicchi and A. Arenas. Abrupt transition in the structural formation of interconnected networks. *Nat. Phys.*, 9(11):717, 2013.

[193] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, P. Chau, K.-P. Chan, S.-V. Lo, P.-Y. Leung, T. Tsang, W. Ho, K.-H. Lee, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, 300(5627):1961, 2003.

[194] S. Rinaldi, J. Peerenboom, and T. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *Control Systems, IEEE*, 21(6):11, 2001.

[195] V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. D. Porcellinis, and R. Setola. Modelling interdependent infrastructures using interacting dynamical models. *Int. J. of Critical Infrastructures*, 4(1):63, 2008.

[196] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy. A long-time limit for world subway networks. *J. R. Soc. Interface*, 2012.

[197] A. F. Rozenfeld, R. Cohen, D. ben Avraham, and S. Havlin. Scale-free networks on lattices. *Phys. Rev. Lett.*, 89(21):218701, 2002.

[198] A. Saumell-Mendiola, M. A. Serrano, and M. Boguñá. Epidemic spreading on interconnected networks. *Phys. Rev. E*, 86(2):026106, 2012.

[199] H. Sayama, I. Pestov, J. Schmidt, B. J. Bush, C. Wong, J. Yamanoi, and T. Gross. Modeling complex systems with adaptive networks. *Computers & Mathematics with Applications*, 65(10):1645, 2013.

[200] W. Schaper and D. Scholz. Factors regulating arteriogenesis. *Arterioscler. Thromb. Vasc. Biol.*, 23(6):1143, 2003.

[201] C. M. Schneider, A. A. Moreira, J. S. Andrade, S. Havlin, and H. J. Herrmann. Mitigation of malicious attacks on networks. *Proc. Natl. Acad. Sci. U.S.A.*, 108(10):3838, 2011.

[202] C. M. Schneider, N. Yazdani, N. A. M. Araújo, S. Havlin, and H. J. Herrmann. Towards designing robust coupled networks. *Sci. Rep.*, 3, 2013.

[203] S. Shai and S. Dobson. Effect of resource constraints on intersimilar coupled networks. *Phys. Rev. E*, 86(6):066120, 2012.

[204] S. Shai and S. Dobson. Coupled adaptive complex networks. *Phys. Rev. E*, 87(4):042812, 2013.

[205] S. Shai, D. Y. Kenett, Y. N. Kenett, M. Faust, S. Dobson, and S. Havlin. Resilience of modular complex networks. *Submitted*, 2014, arXiv:1404.4748.

[206] S. Shao, X. Huang, H. E. Stanley, and S. Havlin. Robustness of a partially interdependent network formed of clustered networks. *Phys. Rev. E*, 89(3):032812, 2014.

[207] L. B. Shaw and I. B. Schwartz. Fluctuating epidemics on adaptive networks. *Phys. Rev. E*, 77(6):066101, 2008.

[208] L. B. Shaw and I. B. Schwartz. Enhanced vaccine control of epidemics in adaptive networks. *Phys. Rev. E*, 81(4):046120, 2010.

[209] L. M. Shekhtman, Y. Berezin, M. M. Danziger, and S. Havlin. Robustness of a network formed of spatially embedded networks. *Phys. Rev. E*, 90(1):012809, 2014.

[210] F. Shi, P.-T. Yap, W. Gao, W. Lin, J. H. Gilmore, and D. Shen. Altered structural connectivity in neonates at genetic risk for schizophrenia: A combined study using morphological and white matter networks. *NeuroImage*, 62(3):1622, 2012.

[211] J. Sienkiewicz and J. A. Hołyst. Statistical analysis of 22 public transport networks in Poland. *Phys. Rev. E*, 72(4):046127, 2005.

[212] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(304):425, 1955.

[213] B. Söderberg. Properties of random graphs with hidden color. *Phys. Rev. E*, 68(2):026107, 2003.

[214] K. Soetaert. *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations*, 2009.

[215] K. Soetaert and P. M. Herman. *A Practical Guide to Ecological Modelling. Using R as a Simulation Platform*. Springer, 2009.

[216] A. Solé-Ribalta, M. De Domenico, N. E. Kouvaris, A. Díaz-Guilera, S. Gómez, and A. Arenas. Spectral properties of the laplacian of multiplex networks. *Phys. Rev. E*, 88(3):032807, 2013.

[217] F. Sorrentino. Synchronization of hypernetworks of coupled dynamical systems. *New J. Phys.*, 14(3):033035, 2012.

[218] V. Spirin and L. A. Mirny. Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U.S.A.*, 100(21):12123, 2003.

[219] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418, 2004.

[220] O. Sporns, C. J. Honey, and R. Kötter. Identification and classification of hubs in brain networks. *PLoS ONE*, 2(10):e1049, 2007.

[221] A. O. Stauffer and V. C. Barbosa. A study of the edge-switching Markov-chain method for the generation of random graphs. *arXiv:cs/0512105*, 2005.

[222] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthélemy. Elementary processes governing the evolution of road networks. *Sci. Rep.*, 2, 2012.

[223] E. Strano, S. Shai, S. Dobson, and M. Barthélemy. Efficiency and centrality of multiplex spatial networks in large urban areas. *Submitted*, 2014.

[224] M. Szell, R. Lambiotte, and S. Thurner. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci. U.S.A.*, 107(31):13636, 2010.

[225] A. Szolnoki and M. Perc. Resolving social dilemmas on evolving random networks. *Europhys. Lett.*, 86(3):30007, 2009.

[226] F. Tan, J. Wu, Y. Xia, and C. K. Tse. Traffic congestion in interconnected complex networks. *Phys. Rev. E*, 89(6):062813, 2014.

[227] F. Tan, Y. Xia, W. Zhang, and X. Jin. Cascading failures of loads in interconnected networks under intentional attack. *Europhys. Lett.*, 102(2):28009, 2013.

[228] P. H. Thrall, J. Antonovics, and D. W. Hall. Host and pathogen coexistence in sexually transmitted and vector-borne diseases characterized by frequency-dependent disease transmission. *The American Naturalist*, 142(3):543, 1993.

[229] L. D. Valdez, P. A. Macri, and L. A. Braunstein. A triple point induced by targeted autonomization on interdependent scale-free networks. *J. Phys. A*, 47(5):055002, 2014.

[230] E. C. van Straaten and C. J. Stam. Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI. *Eur. Neuropsychopharmacol.*, 23(1):7, 2013.

[231] A. Vázquez. Spreading dynamics on heterogeneous populations: Multitype network approach. *Phys. Rev. E*, 74(6):066114, 2006.

[232] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the Internet. *Phys. Rev. E*, 65:066130, 2002.

[233] L. M. Verbrugge. Multiplexity in adult friendships. *Social Forces*, 57(4):1286, 1979.

[234] A. Vespignani. Complex networks: The fragility of interdependency. *Nature*, 464(7291):984, 2010.

[235] M. P. Viana, E. Strano, P. Bordin, and M. Barthélemy. The simplicity of planar networks. *Sci. Rep.*, 3, 2013.

[236] F. Wang, A. Antipova, and S. Porta. Street centrality and land use intensity in Baton Rouge, Louisiana. *J. Transp. Geogr.*, 19(2):285, 2011.

[237] C. P. Warren, L. M. Sander, and I. M. Sokolov. Geography in a scale-free network model. *Phys. Rev. E*, 66(5):056105, 2002.

[238] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[239] S. Watanabe and Y. Kabashima. Cavity-based robustness analysis of interdependent networks: Influences of intranetwork and internetwork degree-degree correlations. *Phys. Rev. E*, 89(1):012808, 2014.

[240] D. J. Watts. *Six Degrees: The Science of a Connected Age*. W. W. Norton & Company, New York, USA, 2003.

[241] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440, 1998.

[242] H. C. White, S. A. Boorman, and R. L. Breiger. Social structure from multiple networks. I. Blockmodels of roles and positions. *Am. J. Sociol.*, 81(4):730, 1976.

[243] S. Wieland, T. Aquino, and A. Nunes. The structure of coevolving infection networks. *Europhys. Lett.*, 100(6):69901, 2012.

[244] H. S. Wilf. *generatingfunctionology*. Academic Press, 1990.

[245] J.-j. Wu, Z.-y. Gao, and H.-j. Sun. Cascade and breakdown in scale-free networks with community structure. *Phys. Rev. E*, 74(6):066111, 2006.

[246] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the Internet's large-scale topology. *Proc. Natl. Acad. Sci. U.S.A.*, 99(21):13382, 2002.

[247] D. H. Zanette. Dynamics of rumor propagation on small-world networks. *Phys. Rev. E*, 65:041908, 2002.

[248] D. Zhao, L. Li, H. Peng, Q. Luo, and Y. Yang. Multiple routes transmitted epidemics on multiplex networks. *Phys. Lett. A*, 378(10):770, 2014.

[249] D. Zhou, J. Gao, H. E. Stanley, and S. Havlin. Percolation of partially interdependent scale-free networks. *Phys. Rev. E*, 87(5):052812, 2013.

[250] D. Zhou, H. E. Stanley, G. D'Agostino, and A. Scala. Assortativity decreases the robustness of interdependent networks. *Phys. Rev. E*, 86(6):066103, 2012.

[251] M. Zimmermann, V. Eguíluz, and M. San Miguel. Coevolution of dynamical states and interactions in dynamic networks. *Phys. Rev. E*, 69(6):065102, 2004.

[252] G. Zschaler and T. Gross. Largenet2: an object-oriented programming library for simulating large adaptive networks. *Bioinformatics*, 29(2):277, 2012.