Brunel University London

Department of Mathematics

Complex Networks with Node Intrinsic Fitness:
On Structural Properties and Contagious Phenomena

Doctoral Dissertation of:
Konrad Hoppe

2014

Abstract

Complex networks is a vibrant research field and has received much attention over the last decade. Central to this area is the question of how networks around us are constructed. The essential notion of network research is that these systems are assembled in a decentralised way, thus no central agent is planning the network beforehand. Despite this lack of central coordination, many networks present intriguing universalities, such as broad degree distributions, in the form of power-laws. The subject of study in this thesis is a class of networks that are constructed by a node intrinsic variable, called fitness. The way these networks grow could be called a rich-get-richer mechanism. The fitter a node is, the more likely it is to acquire new connections inside the network. Several aspects that are directly connected to these networks are explored in this thesis. In the first part, the properties of growing networks that are driven by fitness are investigated and it is shown that the introduction of growth leads to a topological structure that is different from its static counterpart. In the subsequent chapter, percolation on fitness driven networks is studied. The results give insights into possible mechanisms that can stabilise systems. Furthermore, the theory can be used to identify vulnerable structures around us. In the following chapter, the world trade network is discussed. This numerical investigation highlights possible improvements to the methodology to make statistical analysis more robust. That chapter is followed by an analysis of time-varying networks. Time-varying networks represent an interesting construct that allows a formulation of stochastic processes on the same time-scale as the evolution of the network itself. This possibility is highly relevant to the investigation of epidemics, for instance. In the last chapter, a study of a system of clusters and their self-organised formation is presented.

Contents

Chapter 1

Preface

1 Bibliographical Note

Some aspects of this work have been published in:

I. E. Smolyarenko, K. Hoppe, G. J. Rodgers. Network growth model with intrinsic vertex fitness. *Physical Review E* **88**, 012805 (2013).

K. Hoppe, G. J. Rodgers. Mutual selection in time-varying networks. *Physical Review E* **88**, 042804 (2013).

K. Hoppe, G. J. Rodgers. Percolation on fitness-dependent networks with heterogeneous resilience. *Physical Review E* **90**, 012815 (2014).

K. Hoppe, G. J. Rodgers. A microscopic study of the fitness-dependent topology of the world trade network. *Physica A: Statistical Mechanics and its Applications* **419**(1): 64-74 (2015).

2   Acknowledgement

Undertaking this PhD was a great experience and I have enjoyed learning so many new aspects during this time. There are many people without whom this would not have been possible.

First, I would like to thank my supervisor Professor Geoff Rodgers. He has been of such a great support during my entire time of study and always stood behind me during the adverse periods of my PhD curriculum. Moreover, I am very thankful for the great opportunities with which he has provided me. I am very proud to be a student of such a distinguished researcher and I look up to his poise with which he guides any audience and situation.

I also want to thank my parents, Heidelotte Hoppe and Manfred Hoppe who have always loved and supported me. This work would not be possible without them. The process of obtaining a doctorate degree is certainly longer than the core time of study and my parents have prepared me excellently to go this entire path. I am infinitely thankful for all their devotion.

A special acknowledgement goes to my girlfriend Emma Haddi, who was always patient with me and very sensitive to the special circumstances of having a PhD student for a partner (which are plenty). Her love and support made me stay focussed and helped me to refuel during our times together. I love you, Emma.

My thanks also go out to the Department of Mathematics at Brunel that provided me with a generous stipend for the time of my doctoral research. In particular, I want to mention here Kay Lawrence. Her kind and helpful way always made me feel very welcome.

I also want to acknowledge the kind hospitality of Professor Bosiljka Tadic, whom I visited in Ljubljana during my PhD studies. She has given me many interesting insights into the field of dynamical systems and I very much enjoyed our conversations.

Also the staff members of the Graduate School at Brunel University should be mentioned here. I am very thankful for their recognition and support during my time as a doctoral researcher. They have made the experience at Brunel very joyful. I would like to thank especially Diane Smith. She was always very welcoming and helpful and has kindly offered to proofread this work, which is so much appreciated because of my persistent resilience against British rules of punctuation.

Chapter 2

Introduction

In this first chapter, the general concept of networks is introduced. The historical background of network research is illustrated by several examples that are found in the real world. Despite the apparent complex and chaotic structure of many networks, some universal aspects of complex networks are highlighted, followed by various models that can give rise to these universal patterns. The chapter closes with a survey of various types of stochastic processes that can be embedded on complex networks.

1 Networks - The Lingua Franca of Science?

Complex networks is a research field that has drawn knowledge from two directions: the mathematical treatment of graphs and the many particle systems in physics. The social sciences have identified complex networks theory as a viable tool to understand many phenomena around us, such as the spread of rumours, the cascading failure of institutions and the contagion of diseases. Networks are commonly described by a set of nodes and a set of edges that connect the nodes. These sets are also the main objects of study in graph theory. While classical graph theory investigates various properties of graphs that could be relatively small, such as the possibility of different colourings [7], ways of traversing graphs during search [75], or forming round-tours [48], the study of complex

networks that is rooted in physics aims to understand phenomena on almost exclusively very large, and heterogeneous structures. Problems in this area are more probabilistic and need different approaches compared to problems of classical graph theory. For instance, while graph theoretical problems can often be cast in terms of the adjacency matrix of a network, the structure of complex networks is often understood in terms of sets of differential equations. The particular strength of complex networks research lies in the ability to describe local interactions in a much richer way than is possible by using lattices or cellular automata. In particular, the strong heterogeneity of connectivity patterns that appears to be omnipresent in the networks that surround us can easily be modelled using networks, rather than lattices. Also other features, such as the tendency of clustering in social networks, can easily be embedded into networks. In section 4 of this chapter, several constructive models that give rise to large scale heterogeneous structures are presented.

The great interest in complex networks does not merely arise from these points, but rather from the fact that almost everything that surrounds us can be described in terms of a network. Obvious examples are the World Wide Web [4, 70] that is spanned by webpages, connected with hyperlinks. Also the physical internet that is spanned by routers and cables has been investigated [52]. Other examples include the network of movie actors [8], where nodes are actors and edges connect those that have acted in a movie together. Similarly, scientific co-authorship networks have been investigated [8, 9, 90], where nodes represent individual authors and edges connect academics that have co-authored papers together. Other examples are transportation networks [12, 35, 63] and the power-grid [2, 36]. More interesting than the mere description of the networks is the analysis of processes that unravel on these. For instance, the prediction of disease epidemics [35, 83, 120], and the spread of memes and information diffusion in social networks [29, 126]. Furthermore, biological networks such as the structure and function of the brain have been investigated [23] as well as protein encoding networks to better understand the microscopic structure of our bodies and human diseases [118]. This incomplete list of possible applications of complex networks research provides a glance of the strength of this academic discipline. However, the application of networks is not

limited to these large scale structures. As initially mentioned, mathematical graph theory can be utilised to solve a variety of problems, such as colouring and routing. Other applications of networks can be found in machine learning where artificial neural networks are employed to mimic the human brain and *understand* complex data [106]. Finally, computers perform elementary algebra by traversing trees which are a special class of networks. The same structure is used for algorithms in mathematical logic. Thus, even theorems can be proved using networks.

In summary, networks appear in a multitude of scientific disciplines and many problems can be translated into graphs. The more recent history of complex networks, which are classical graphs with the additional feature of being large and heterogeneous, has proven extremely successful in tackling problems of the present century [65].

2 Examples of Networks

2.1 Seven Bridges of Königsberg

Problems related to graphs were explored early in the history of mathematics. A very famous problem is the Seven Bridges of Königsberg. Fig. 2.1a shows a map of Königsberg. The task is to find a round tour through the city that visits every bridge exactly once. Leonhard Euler proved in 1736 that this path does not exist [48]. Although the problem is embedded in a geographic setting, its spatial dimension is irrelevant, because the position of the bridges does not matter. The only aspects that matter are the connecting bridges. Therefore the problem can be represented as a graph, with the bridges as edges. Fig. 2.1b illustrates this representation. Euler's article is regarded as one of the earliest in the field of graph theory and, the problem of finding an Eulerian tour has become a part of standard theory.
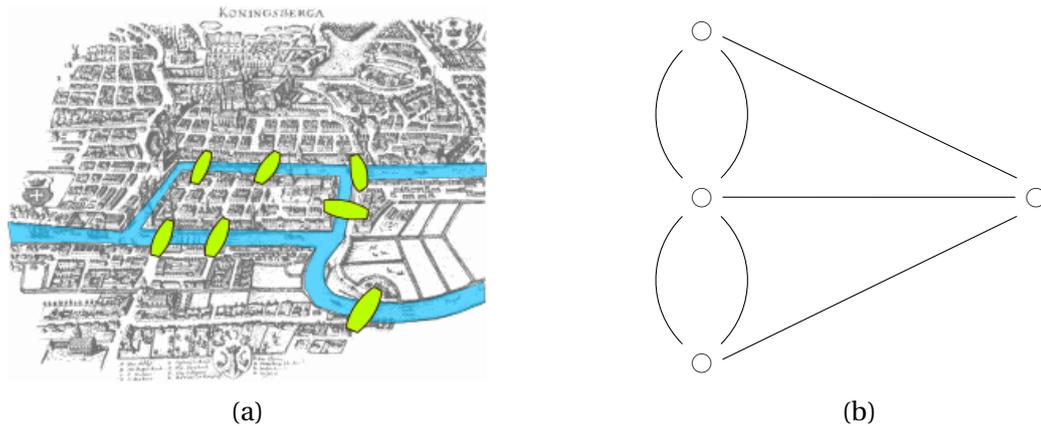
Figure 2.1: (a)An ancient map of Königsberg with the river system in blue and the seven bridges marked in green. (b) Graphical representation of the bridges.

2.2 The Medici

The House of Medici was a very influential family in Florence that brought forth two queens during the 15th and 16th centuries, as well as four popes in the course of the 15th-17th centuries. The family came to fame and political power for the first time under Cosimo de' Medici during the 14th century in Florence. The key to their sudden prosperity was not their political dominance or wealth at that time, but rather their central position in the network of marriages between different families [95]. Marriages between families have played a key role in arranging business deals and forming political allies. Fig. 2.2 illustrates this network of inter-marriages. It is obvious from the figure that the Medici inhabit a very central position in this network. This central position was their advantage: many consolidations – commercial or political – between the families of the Italian Renaissance were intermediated by the Medici. This graphical indicator of importance in the network can also be quantified. Betweenness centrality of a node measures the fraction of all shortest paths from all vertices to all others that pass through that node. To make this more clear, define $s_{ij}(k)$ as the number of shortest paths that go from node i to node j through node k and s_{ij} as the total number of shortest paths from node i to j. For instance, the number of shortest paths from Guadagni to Salviati is 2, and so is the number of paths that route via the Medici. Thus $s_{\text{Guadagni, Salviati}} = 2$, and $s_{\text{Guadagni, Salviati}}(\text{Medici}) = 2$, while $s_{\text{Guadagni, Salviati}}(\text{Albizzi}) = s_{\text{Guadagni, Salviati}}(\text{Tornabuon}) =$
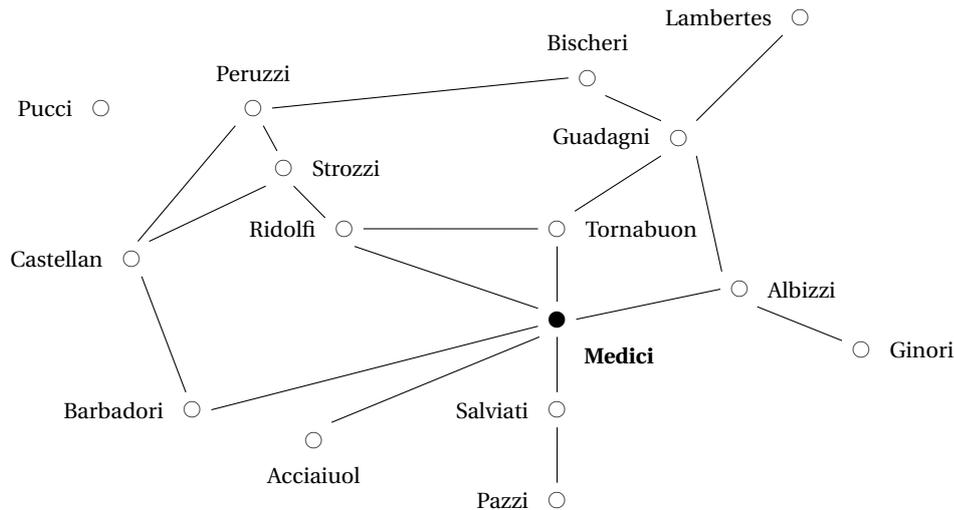
Figure 2.2: Network of Florentine marriages in the 15th century. Nodes mark the families and edges marriages between the members of them.

1. The overall betweenness of a node is then

$$c(k) = \frac{1}{(N-1)(N-2)/2} \sum_{i \neq k \neq j} \frac{s_{ij}(k)}{s_{ij}}.$$

(2.1)

The factor $2/(N-1)(N-2)$ normalises $c(k)$ to the unit interval. Interestingly the Medici score the highest in terms of betweenness centrality with $c(k) = 0.522$, while the next highest centrality is found for the Guadagni with $c(k) = 0.255$.

Another measure of importance is the number of adjacent edges of a node. This number is called the node degree. Also in terms of the node degree, the Medici score top with six adjacent edges, followed by the Strozzi and Guadagni with four adjacent edges each.

Given the subsequent rise of the Medici's power in the 15th and 16th centuries, it is clear that an investigation of the network structure of Florentine marriages provides viable insights into future outcomes that lay beyond plain marriage patterns. The next subsection will give some insights into a more recent network of interest, the network of terrorists who were involved in the 9/11 attacks. Also in that case, simple "back of the envelope calculations" give useful insights into aspects that have later been identified by more thorough investigations.

2.3 The 9/11 Terrorists' Network

Shortly after the 9/11 attacks on the USA, information on the social ties of the terrorists became publicly available. Nowadays, it is commonly known that Mohamed Atta was one of the central planners of the attack. Fig. 2.3 illustrates the network of acquaintances between the plane hijackers of 9/11. From that figure, it is clear that Mohammed Atta not
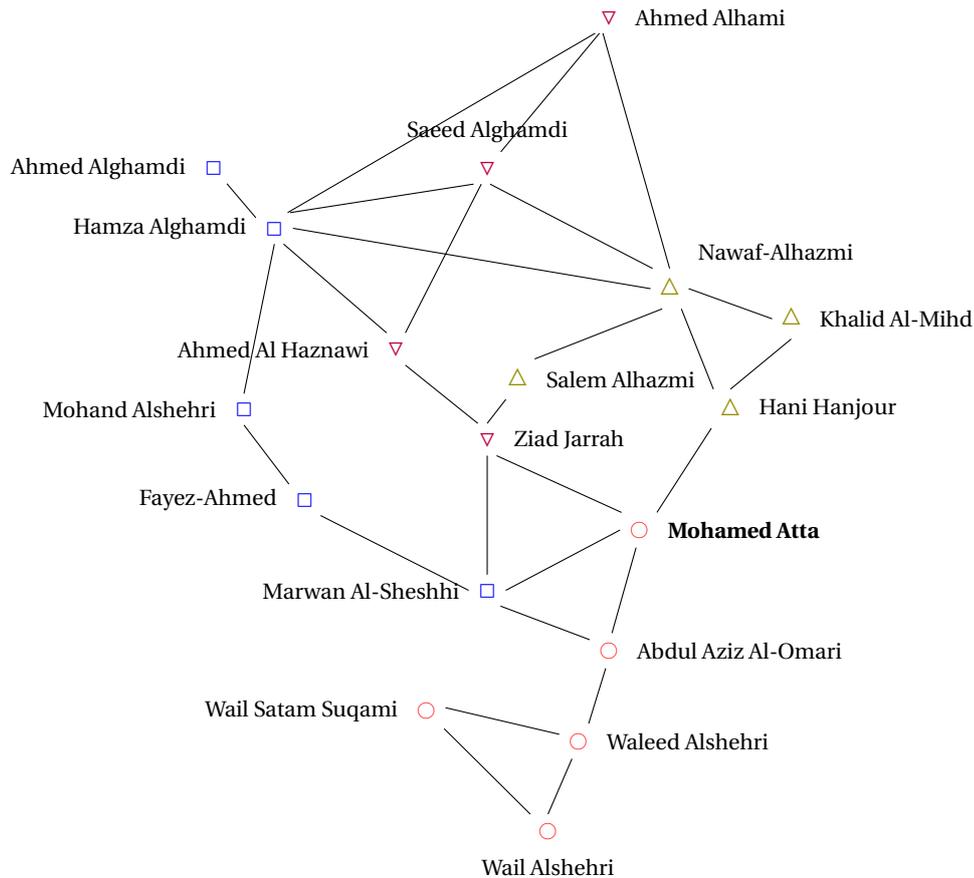
Figure 2.3: Network of terrorists involved in the 9/11 attacks. ○: Flight AA #11 Crashed into the World Trade Centre. □: Flight UA#175 Crashed into WTC. △: Flight AA #77 Crashed into Pentagon. ▽: Flight UA #93 crashed in Pennsylvania.

only has a very central position within the network, but also serves as a intermediator between the different teams that hijacked the four airplanes. Krebs [76] shows a larger snapshot of this network with links also to other accomplices who provided money and logistics. Considering this larger network, Mohamed Atta reserves a central role with a betweenness centrality of 0.588 followed by Essid Sami Ben Khemais with a value of 0.252.
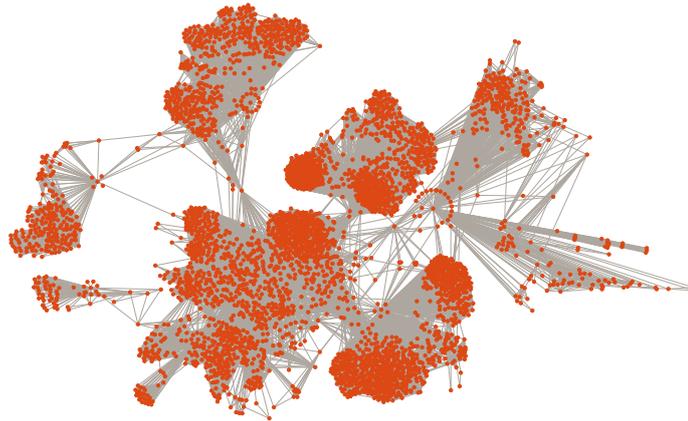
Figure 2.4: Snapshot of the network of friendships on Facebook (source: Stanford Large Network Dataset Collection)

This is therefore another example of the importance of network metrics that shed light on aspects far beyond the scope of topological insights. The network of terrorists and further consequences of its topology have also been investigated in [99]. In particular, it is important to note that an understanding of the underlying network/hierarchy structure is essential for a successful anti-terror operation. The impact of different network structures on deliberate attempts to disturb the function of a network is discussed later in this chapter.

3 Universal Structures of Complex Networks

The networks around us are of astonishing size and complexity, see Fig. 2.4 for an example of a snapshot of Facebook. However, there are also a number of interesting regularities in these structures. To illustrate this concept, some features of three different networks are investigated in this section: a graph of the world wide web provided by Google, the collaboration network in the area of condensed matter on arXiv[1], and the network of publicly shared circles on Google+[2]. All data has been obtained from the Stanford Large Network Dataset Collection[3].

Many networks are so called small-world networks. These networks are characterised in

---

[1] www.arxiv.org
[2] https://plus.google.com
[3] https://snap.stanford.edu/data/

terms of their average path length L. That is the number of steps it takes on average to walk from one node to another, whereby steps can only be made along edges. A network is identified as a small world if the average path length growths logarithmically with its size [125], thus

$$L \propto \log N, \tag{2.2}$$

where N is the number of nodes. The small-world property is sometimes also known as the six degrees of separation, coined by Guare [62]. Another property of interest is the local clustering coefficient c_i of a node. The clustering coefficient measures how close the neighbours of a node are to being a clique. A clique is a set of nodes that is fully connected, such that every possible edge exists. Formally, the local clustering coefficient is defined as

$$c_i = \frac{1}{k_i(k_i - 1)} \left| \left\{ e_{jk} : v_j, v_k \in \mathcal{N}_i, e_{jk} \in E \right\} \right| \tag{2.3}$$

where $\mathcal{N}_i$ is the set of vertices in the direct neighbourhood of node i. Furthermore, $E = \{e_{i,j} : i, j = 1, \dots, N\}$ denotes the set of all edges and V the set of all nodes with $V = \{v_i : i = 1, \dots, N\}$. For an undirected network, the concept can be accordingly defined, but needs renormalisation by factor 2. Fig. 2.5 illustrates the local clustering coefficient graphically. The average clustering coefficient is then given by

| (a) | (b) | (c) |
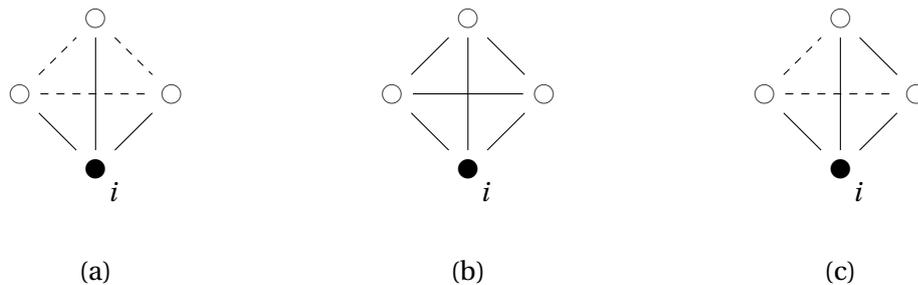
Figure 2.5: Example configurations of a node's neighbourhood. Solid edges mark present edges, dashed edges possible ones. The local clustering coefficient is for the different configurations: (a) $c_i = 0$, (b) $c_i = 1$, and (c) $c_i = 1/3$.

$$\langle c \rangle = \sum_{i=1}^{N} \frac{c_i}{N}. \tag{2.4}$$

Real-world networks show significantly more clustering than random networks. The most simple random network model was proposed by Erdős and Rényi [46] and serves

as a null model for many network concepts. The model is described by a fixed number of nodes and edges. Each edge exists with probability p, such that the existence of edges are statistically independent from each other. In this case the clustering coefficient is simply $\langle c \rangle = p$, since the probability that pairs of edges exist in the neighbourhood does not depend on any neighbourhood linked property. Clustering describes the cliquishness of a network. In terms of friendship networks, high clustering implies that the friends of someone are themselves friends of each other. This transitivity of friendships is likely in social networks and thus, real-world networks exhibit much higher clustering than pure random graphs. Tab. 2.1 summarises these network measures for three different networks. Another feature that can be observed in real-world networks is that they have

	N	M	$\langle c \rangle$	Diameter
Google web graph	$875,713$	$5,105,039$	0.5413	21
arXiv cond-mat	$23,133$	93497	0.6334	14
Google+ circles	$107,614$	13673453	0.4901	6

Table 2.1: Central properties of the three networks under investigation, with N as the number of nodes, M the number of edges, $\langle c \rangle$ the average clustering coefficient, and the diameter as the longest shortest path in the network.

broad degree distributions, to be more precise: power-law degree distributions [31]. The degree distribution $p(k)$ is the probability that a randomly chosen node has k neighbours. In the case of a power-law, $p(k)$ has the form

$$p(k) \propto k^{-a}. \tag{2.5}$$

Power-laws appear in many areas of interaction, also in the three networks that serve as examples in this section; see Fig. 2.6. It is not entirely clear why interaction leads regularly to power-law distributed properties. In fact there is an entire branch of literature investigating this phenomenon [84]. For networks, one mechanism that is believed to drive networks into the observed power-law degree distributions is preferential attachment [8]. The notion of preferential attachment is that interactors with more existing interactions are more likely to engage in new interactions. In terms of the network of friendships, a person with many friends finds it easier to acquire new relationships, than somebody who is entirely isolated. More details on this model can be found in the following section.
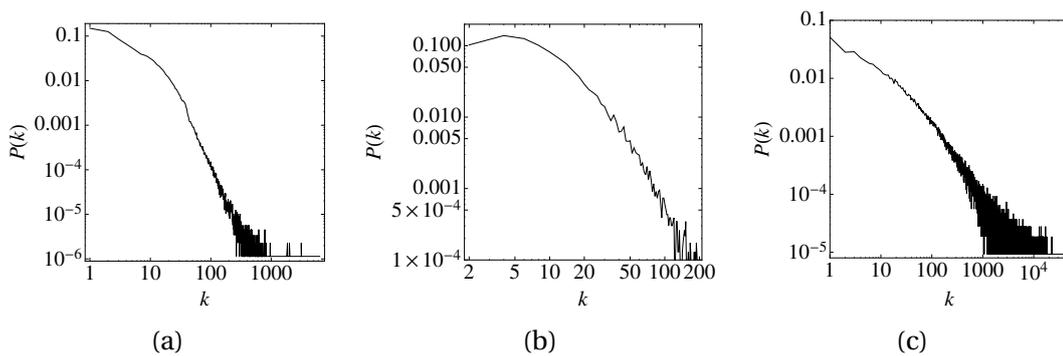
Figure 2.6: Degree distributions of (a) the web graph (b) the collaboration network in the condensed matter section on arXiv (c) the friendship network on Google+.

4   Models of Networks

In section 2, some examples of the usefulness of network studies have been illustrated. While the focus was put on the existence of various observable artefacts, in this section a number of models that lead to certain macroscopic patterns will be reviewed.

One of the first and simplest network models is the Erdös-Rényi (ER) Model [46]. Two variants of this model exist: the $G(N, M)$ model, that describes an undirected graph with N nodes and M edges, and the $G(N, p)$ model, which elucidates a graph with N nodes whereby each pair of nodes shares an edge with probability p. In both cases the number of nodes is static and edges are deployed randomly. The resulting degree distribution is binomial and can be approximated in the $N \to \infty$ regime as a Poisson distribution

$$P_{\mathrm{ER}}(k) = \frac{(Np)^k e^{-Np}}{k!} \tag{2.6}$$

for the $G(N, p)$ model. The relationship to the $G(N, M)$ model is then described by $M = \binom{N}{2}p$.

The ER model is a useful null model to study networks, but it lacks several features that are observed in real-world networks. One such feature is clustering. Networks around us are highly clustered and have short average path lengths. A model that resembles this feature is the Watts-Strogatz (WS) model [125]. The network consists of a fixed number of nodes. All nodes are organised on a ring lattice. The central parameter of the model

are N, k^* and p. N is the number of nodes, k^* the number of nearest neighbours on the ring lattice that each node is connected to and p is the probability that an existing edge gets rewired. p drives the structure of the network from a regular ring lattice for $p = 0$ via a small-world network for $0 < p < 1$ towards a random network at $p = 1$. The rewiring takes place in a sequential process that considers each edge exactly once. More details can be found in [125]. The degree distribution depends on the value of p. For $p = 0$, $p(k)$ is peaked at k^*: $p(k) = \delta(k - k^*)$ where $\delta(x)$ is the Dirac delta function. For $p = 1$, the degree distribution is Poissonian, as in the ER model. Also for values of $0 < p < 1$, the degree distribution can be computed in terms of a series representation [14].

A further generalisation of the Erdös-Rényi model has been introduced in [24]. The model by Caldarelli et. al. extends the ER model by introducing a vertex intrinsic variable and an attachment kernel $f(x, y)$, which is defined as the probability that a pair of nodes with intrinsic variables x and y share an edge in a network with N nodes. The hidden variable is sometimes called fitness, which refers to its ability to attract edges inside the network. The degree distribution $p(k)$ can be found modelling a more detailed quantity, namely $p(k|x)$, the fitness conditional degree distribution. One possible approach is to expand $p(k|x)$ into its contributions [21, Eq. (15)]. Another approach is to derive it from first principles, noticing that the static nature of the network can be ignored for the sake of the argument and edges deployed sequentially. In this case the evolution of the fitness conditional degree distribution in a network with M edges and N nodes can be described by

$$p_{M+1,N}(k|x) = p_{M,N}(k|x)[1 - \lambda(x, N)] + p_{M,N}(k - 1|x)\lambda(x, N), \qquad (2.7)$$

where $\lambda(x, N)$ is the probability that a node with fitness x receives an edge during an edge deployment step. The details of this derivation are discussed later in Chapter 5, following Eq. (5.5). The resulting fitness dependent node degree distribution is found equivalent to [21] to be Poissonian:

$$p_{M,N}(k|x) = \frac{e^{-\bar{k}(x)}[\bar{k}(x)]^k}{\Gamma(k + 1)}, \qquad (2.8)$$

where $\bar{k}(x)$ is the fitness conditional degree expectation. The exact form of the degree

distribution of the resulting network depends then on the probability distribution of fitness and the attachment kernel $f(x,y)$. The fitness model has also been analysed in [111], where the inverse and direct problems have been addressed. That is the problem of finding analytical forms of $f(x,y)$ and the fitness distribution $\rho(x)$, such that the resulting network exhibits a given degree distribution. The solutions to these problems are of high importance when the internal mechanisms of the network need to be understood in more detail. The spreading of diseases can usually be expressed in terms of $f(x,y)$ and $\rho(x)$, and solutions to the inverse and direct problem can give rise to these quantities.

In Chapter 3, a further extension of the fitness model, introducing growth is studied. Growth has been proven to be an important ingredient for many real-world models [42]. The fitness conditional degree distribution can be found with the same technique as outlined above for the static case. This distribution is given by

$$p(k|x) = \frac{1}{1 + \lambda(x)} \left(\frac{\lambda(x)}{1 + \lambda(x)} \right)^{k-1}.$$ (2.9)

Whereby $\lambda(x)$ is similarly defined to that in the previous paragraph on static fitness driven networks. More details of this model can be found in Chapter 3.

Another class of network models that have attracted significant amount of attention in the complex networks literature are degree based models. Early work on this class was motivated by bibliographic studies of cross-referencing between different publications [102]. The cumulative advantage of having more citations in order to receive more following citations in the future was studied in more detail in [103]. However, Price's investigations were not widely recognised until 1999. Barabási and Albert [8] reinvestigated what is nowadays called the Barabási Albert (BA) model. The BA model is a growing network model, that models the cumulative advantage by imposing that each new node that arrives to the network connects preferentially to nodes with high degree. More formally, define the probability that a new node connects to an existing node i as

$$\pi_{\text{new} \to i} = \frac{k_i}{\sum_j k_j} = \frac{k_i}{2t}.$$ (2.10)

k_i is the degree of node i. Thus, the rate at which the degree of a node changes is

$$\frac{d k_i}{d t} = k_i/2t.$$ (2.11)

Therefore, the final degree distribution is then found to be

$$p(k) = 2k^{-3}.$$ (2.12)

This finding is the reason for the success of the BA model. Many real-world networks exhibit power-law degree distributions [3]. Interestingly both features, growth and preferential attachment, have to be present. Growth without preferential attachment leads to an exponential distribution, while preferential attachment without growth leads to a non-stationary distribution, ultimately resembling a complete graph, that is a network with all possible edges present [8].

Next to the two classes of models that are built either exclusively by a hidden variable or exclusively by a degree based mechanism, there exists also a third, a hybrid type. Bianconi and Barabási [19] introduce a model which is motivated by the finding that in the classical BA model old nodes had more chances to connect than young nodes and will therefore have on average a higher degree. In real-world networks one can frequently find nodes that arrive new to the network and are able to acquire large numbers of new links in a short time. These out-performers must have some intrinsic quality which drives this rapid growth. To account for this, a generalised attachment kernel that considers this increased ability is introduced:

$$\pi_{\text{new}\to i} \frac{k_i x_i}{\sum_{j=1}^{N(t)} k_j x_j},$$ (2.13)

where x_i is the fitness of node i and k_i its degree. A generalisation of this model with other approaches to link degree and fitness was introduced by Ergün and Rodgers [47]. Another form of combining the fitness model and the degree preferential attachment has been investigated in [16]. Bedogne and Rodgers [16] introduce a model where one of two things can happen at each time step. Either a new node with one link is added, whereby

the target node is chosen degree preferentially as in the BA model, or a new edge between two existing nodes is deployed, depending on the nodes' fitness. The resulting degree distribution is a pure power-law. The exponent depends only on the relative frequency of node additions compared to edge addition steps.

Perra et al. [100] present a model that is characterised by dynamics on a much shorter time scale than in the previously discussed models. The preceding models are either entirely static or growth occurs gradually, one update at a time. The time-varying network model in [100] is entirely different. Time varying networks are described by a fixed number of nodes and a varying number of edges. If the model is described in discrete time, then at each time step each node becomes active with a node intrinsic probability and connects to a fixed number of existing nodes inside the network. The motivation behind this class of models is, for example, in the mobile phone network, where nodes represent mobile phones and edges phone calls. If a phone initiates a call to another, then these two nodes share an edge during that particular instance. Other examples include human contact to spread diseases or even rumours. The particular appeal of a time-varying network model lays in the possibility to formulate the evolution of the network on the same timescale as the unfolding of stochastic processes on top of it. More details about this class of networks can be found in Chapter 6.

The preceding network models are flat, in the sense that edges are just of one type and binary. However, there exist a number of systems for which these assumptions do not hold. Investigating the structure of link weights is important in a number of applications. Consider for instance the strength of a social tie, distinguishing long time friends from random encounters. Another example forms the air-route network, where nodes are airports and edges the flight routes between them. The weight of a link might describe for instance the number of seats that are available during a day on a given route. This knowledge can be used to comprehend the contagion of infectious diseases or the severity of service interruptions. Additionally, edges cannot just have different weights, but can also be of different types. Continuing the airline network example from above, cities are not only connected by air-routes, but also by streets and the railway system. These different layers can either be represented as separate networks or as different types

of links. Fig. 2.7 illustrates this notion. Also the network of international trade forms an
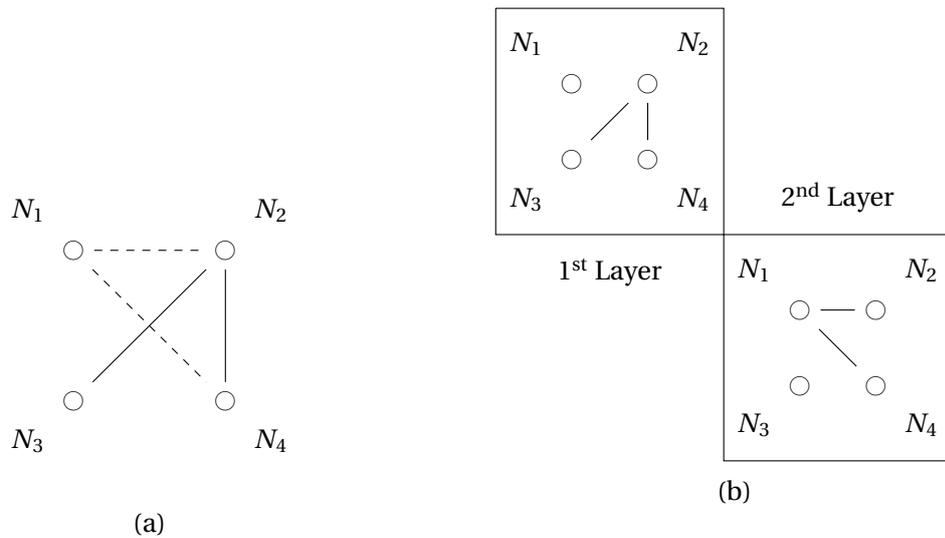
Figure 2.7: Schematic illustration of multiplex networks. (a) Representation of multiplicity as different edge types. (b) Representation of multiplicity as different network layers.

example of a multiplex network. Each node represents a country and the different edge types indicate the traded amount of single product classes between pairs of countries. Another example is the power supply network, that consists of different levels, such as gas pipelines, high voltage power lines, water pipes, etc. Apart from merely describing these structures, it is of central interest to understand how to prevent failures of these networks and protect them against deliberate attacks. The study of these phenomena is the subject of the following section.

5 Stochastic Processes on Networks

Many real-world phenomena can be described in terms of networks, such as trade relations, human contacts, collaborations, etc. Beyond a detailed description of these networks, it is of great importance to understand how these networks behave in the presence of adverse events and to what extent their function is compromised. In this section, a number of stochastic processes that occur on networks are discussed.

5.1   Targeted Attacks and Random Failures

Random networks described in the ER model and scale-free networks produced by the BA model are very different with respect to their topological quantities. These differences prove to be crucial when the networks are under targeted attack, or subject to random failures. Albert, Jeong, and Barabási [5] present a numerical investigation of the consequences for both type of networks. The scale-free (SF) network is very resilient against random failures compared to the random network. The reason is the heterogeneous contact pattern of the SF network. There is just a small number of hubs, that connect large proportions of the network. Random failures are unlikely to occur at these hubs, so a large fraction of the network can fail before the network's function is affected. The behaviour of the random network is entirely different. Most nodes have degree $k \sim \langle k \rangle$, thus each node is equally important for the function of the network, and hence removing nodes increases the diameter of the network linearly.

While the SF network is extremely resilient against random failures, it is prone to fail quickly under targeted attack. An attacker who knows the hubs of the network that connect large proportions can destroy the network's function after a small number of removals. In this case, the random network has the advantage. Since every node connects on average the same fraction in the network, the additional knowledge of the attacker does not contain any valuable information and thus the diameter increases in the same way as in the case of random failures.

5.2   Epidemic Spreading

The study of epidemics has received attention from the mathematical community for a long time. One of the earliest contributions has been made by the Swiss mathematician Daniel Bernoulli. He investigated a mathematical model of smallpox in 1760 [17]. A comprehensive review of the coverage of infectious diseases in mathematics can be found for instance in [66].

Before an underlying network structure was considered, the diseased population was assumed to be perfectly mixed in the sense that an individual has equal probabilities to connect to any other individual in the population to pass on or to receive the disease. This assumption allows one to describe the evolution of an epidemic in terms of simple linear differential equations. A simple compartmental epidemic model is the susceptible-infected-susceptible (SIS) model. In this model each node belongs to one of the two compartments: infected I or susceptible S. If the population is assumed to be perfectly mixed, then the evolution of the number of individuals in each compartment is governed by

$$\frac{ds}{dt} = -\beta s i + \gamma i \quad \text{and} \tag{2.14}$$

$$\frac{di}{dt} = \beta s i - \gamma i. \tag{2.15}$$

where $s = S/N$ and $i = I/N$ are the fractions of susceptible, respectively infected individuals. β is the probability that a susceptible individual gets infected in a contact with an infected individual and $1/\gamma$ is the time it takes to recover from the disease. Fig. 2.8 illustrates the development of the size of these two groups.
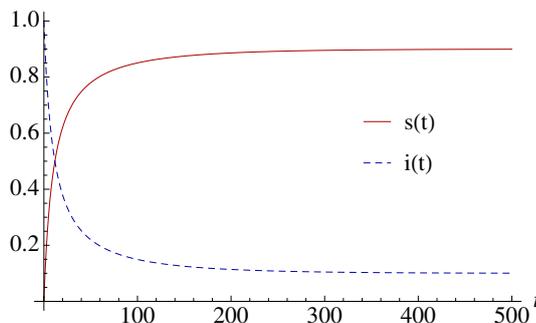
Figure 2.8: Size of the two compartments: susceptible (S) and infected (I) in the SIS model with $\beta = 1/10$ and $\gamma = 9/100$.

The figure shows that the number of individuals in each compartment reaches an equilibrium value after a relatively short burn-in phase. A central quantity of epidemic research is the reproduction number that measures the number of secondary infections following a single infectious individual being introduced to the system. This number is

defined as

$$R_0 = \beta/\gamma. \tag{2.16}$$

R_0 describes the potential of the disease. In particular, if $R_0 < 1$, the infectious individual cannot generate an epidemic, since there are not enough secondary cases in order to sustain the disease. If however $R_0 > 1$, then the diseased population grows *ad infinitum*. If $R_0 = 1$, the disease will stay forever in the population, since every single case produces one successive case. $R_0 = 1$ marks the epidemic threshold that separates the two phases of infinite growth of the infectious fraction and instantaneous immunity.

The preceding technology can easily be expanded to account for networks in which the degree distribution is peaked at $\langle k \rangle$ and decays exponentially [13]. For heterogeneous networks, this is not the case. A typical node degree does not exist as it is possible to find many nodes with $k \gg \langle k \rangle$. It is therefore necessary to consider the different degree classes separately, to compute the evolution of the epidemics. The variables of interest are the fraction of infected nodes with degree k: $i_k(t) = I_k/N_k$, and the fraction of susceptible individuals in degree class k $s_k(t) = S_k/N_k$. The evolution of the SIS model is then described by

$$\frac{d i_k}{d t} = \beta k s_k \Theta_k(t) - \gamma i_k \quad \text{and} \tag{2.17}$$

$$s_k = 1 - i_k. \tag{2.18}$$

Θ_k is the density of infected neighbours around a node with degree k:

$$\Theta_k(t) = \sum_{k'} P(k'|k) i_{k'}(t) \tag{2.19}$$

In the absence of degree correlations, this quality becomes

$$\Theta_k(t) = \Theta(t) = \frac{1}{\langle k \rangle} \sum_{k'} k' P(k') i_{k'}(t). \tag{2.20}$$

Notice, that the k independence holds only in the absence of degree correlations. Without degree correlations the conditional probability to find a node with degree k' in the

neighbourhood of a node with degree k is given by

$$p(k|k') = \frac{k\,P(k)}{\langle k \rangle}. \tag{2.21}$$

The epidemic threshold is found to be

$$\frac{\beta}{\gamma} \geq \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{2.22}$$

It has been illustrated that for many real-world networks that the degree distribution follows a power-law $p(k) \propto k^{-a}$ with a typical exponent $2 < a < 3$. Thus in the $N \to \infty$ limit, $\langle k^2 \rangle \to \infty$ and therefore an epidemic threshold of zero is found, which implies that a disease can never die out on a heterogeneous network. Real-world networks have finite size and therefore a non-zero epidemic threshold can be computed, but nevertheless this result shows that epidemics can spread over a heterogeneous network easily even with very low transmission rates.

5.3 Opinion Formation

Opinion or consensus formation is a widely studied phenomenon in social sciences and has received attention in the complex networks literature as well. The simplest models are the voter model (VM), the invasion model (IM) and the link dynamics model (LD). All these models comprise a set of agents that are characterised by a discrete, usually binary variable $s_i = \pm 1$ that reflects their opinion on a certain topic, for instance Republican vs. Democrat. Independent of the embedding geometry, the evolution of these three models is usually described in discrete time. In the voter model, at each time step a node is chosen randomly and adopts the opinion of a randomly chosen neighbour. In the invasion model, the randomly picked neighbour adopts the opinion of the initially chosen node. In the LD model, an edge is chosen randomly and a randomly chosen node adjacent to this edge adopts the opinion of the node on the other side of that edge. Early work on voting and invasion can be found in [32]. If the community of agents is embedded on a d-dimensional lattice, then the community reaches an ordered state, that is a

situation where each agent has the same opinion, for $d = 1, 2$. Castellano, Vilone, and Vespignani [26] derive results for voting on the small-world network (WS model) [125]. Interestingly, long range interaction induced by the rewiring in the WS model leads to a faster ordering/agreement of the system only in finitely sized networks. In the thermodynamic limit of $N \to \infty$, the small-world network reaches a non-trivial equilibrium, in which both opinions can coexist. Sood and Redner [113] investigate the time it takes until a consensus on a heterogeneous network can be reached. This time will be called consensus time T_N. The consensus time on a regular d-dimensional lattice scales with as $T_N \propto N^2$ for $d = 1$ and $T_N \propto N \ln N$ for $d = 2$ and $T_N \propto N$ for $d > 2$ [113]. This scaling behaviour is found to be different on heterogeneous graphs. For power-law networks for example, the scaling is found to be for $p(k) \propto k^{-a}$ as $T_N \propto N \ln N$ for $a = 3$, $N^{(2a-4)/(a-1)}$ for $2 < a < 3$ and $T_N \propto (\ln N)^2$ for $a = 2$ [113]. More results on the invasion model and the link dynamics model can be found in [112].

5.4   Percolation Phenomena on Complex Networks

Percolation theory on complex networks is concerned with topological properties of occupying clusters on top of a network. Occupation of a single node can be understood as a diseased state and percolation theory finds therefore applications in epidemic modelling [91]. The topological properties of the percolating component that is described by occupied sites and links between those can be understood in three distinct phases. These phases are separated by the percolation threshold p_c. For low occupation probabilities $p < p_c$, many small clusters exist. If the occupation probability is greater than the threshold, i.e. $p > p_c$, then the network is occupied by one giant component that is globally connected. In terms of an epidemic, this state describes the situation when a disease has broken out and a major part of the population is infected. At $p = p_c$, the system is critical and occupied clusters of all sizes exist.

To compute the properties of the giant component, define q as the probability that a randomly chosen edge does not lead to the percolating component. This probability q can be calculated self consistently. Consider randomly picking a node i_0 and follow

one of its links randomly toward a node i_1. The link between i_0 and i_1 is only not part of the percolating component if none of the other links attached to i_1 are not part of the component. The overall probability for this event can be split into two parts: the probability that the randomly picked link ends at a node with degree k and that none of the other $k-1$ links belongs to the percolating component. In an uncorrelated network, $p(k|k')$ is independent of k' and the probability that none of the $k-1$ links belongs to the percolating component is just q^{k-1}. Therefore q can be computed by

$$q = \sum_k \frac{k p(k)}{\langle k \rangle} q^{k-1}. \tag{2.23}$$

The construction above appears cumbersome at first glance, because of its construction in terms of probabilities that a event does not occur. However, the readers can convince themselves that the positive formulation would demand a more detailed analytic expression.

Define now the probability that a node belongs to the giant component as p_G. Obviously,

$$1 - p_G = \sum_k p(k) q^k. \tag{2.24}$$

This probability can be computed by solving Eq. (2.23) and substituting for q. A trivial solution to Eq. (2.23) is $q = 1$. Substituting $q = 1$ into Eq. (2.24) leads to $p_G = 0$. There exists however maximally one more non-trivial solution to Eq. (2.23). Define $G(q) = \sum_k \frac{k p(k)}{\langle k \rangle} q^{k-1}$. Thus Eq. (2.23) can be rewritten to $G(q) = q$. Notice that $G(0) = p(1)/\langle k \rangle$ and $G(1) = 1$. Further is $G'(q) > 0$ and also $G''(q) > 0$. Thus $G(q)$ is monotonously increasing and convex. That also implies that $G(q) = q$ can have maximally one more solution than $q = 1$. Fig. 2.9 illustrates the only two possible configurations. More formally, $G(q) = q$ can only have a second solution if the slope of $G(q)$ is higher at $q = 1$ than the slope of q:

$$\left. \frac{d}{dq} G(q) \right|_{q=1} > 1. \tag{2.25}$$

Substituting for $G(q)$ leads to

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2. \tag{2.26}$$
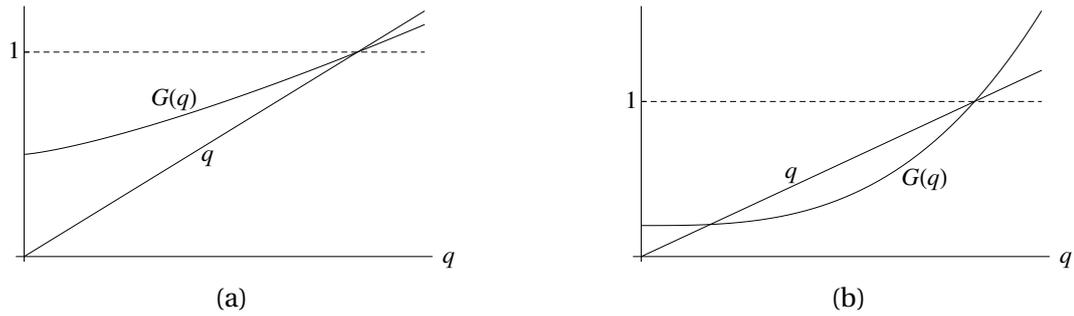
Figure 2.9: Illustration of Eq. (2.23). (a) Only the trivial solution to $G(q) = q$ at $q = 1$ exists. (b) An additional solution $0 < q < 1$ exists.

This equation marks the percolation threshold of a network with arbitrary degree distribution, without degree correlations. More details on this derivation and information on how to calculate related quantities can be found for example in [13, 34, 85, 93]. Cohen et al. [34] use this technology to investigate how many nodes need to be removed from the internet until a complete breakdown of the system, in the sense that a failed component reaches global connectivity. Since the internet's degree distribution follows a power-law $p(k) \propto k^{-a}$ with $a \approx 5/2$ [52], the result in Eq. (2.26) can be used to show that the internet is extremely resilient, in fact it still functions after 99% of the nodes have been removed [34].

Chapter 3

A Network Growth Model with Node Intrinsic Fitness

In the previous chapter, the main concepts of networks were introduced alongside some motivating examples showing how network science is relevant to many disciplines. In this chapter, a particular class of networks is introduced and analysed. The class of fitness driven, growing networks. This class of models is classically characterised by a fitness linking mechanism that governs the attachment rate of new links to existing nodes, and a distribution of node fitness that measures the attractiveness of a node. At each time step of this dynamical model, either a new node joins the network and is attached to one of the existing nodes, or a new edge is added between two existing nodes with probability proportional to the nodes' attractiveness. The full analytic theory connecting the fitness distribution, the linking function, and the degree distribution is constructed. Given any two of these characteristics, the third one can be directly determined. One of the results is that the ubiquitous power-law degree distributions that were introduced in the previous chapter require very specific conditions on the attachment kernel or fitness distribution.

1   Introduction

Numerous problems arising in the social and natural sciences (biology, sociology, information systems, etc.) can be cast in terms of networks of interacting agents. Network theory, built on the basis of graph theory, has become a powerful tool for the study of such systems in the last dozen years [3, 20, 89]. In such descriptions, agents are represented as nodes of a graph, while edges encode interactions. As it has been argued in the previous chapter, many of these networks show common topological features [3]. In particular it has been found that many of those real-world networks have low average path lengths and exhibit power-law degree distributions. This class of networks are called scale-free networks [42].

A model that resembles these two features very well is the Barabási-Albert (BA) Model of network growth. Two essential components lead to this particular network topology: growth and preferential attachment. Both of these mechanisms are important, each acting in isolation do not yield small-world networks with power-law degree distribution [8].

But while the BA model of network growth is theoretically a very robust construction, that leads naturally to power-law degree distributions, it has been found that degree-preferential attachment is not the process which is empirically evident on growing networks [72]. Jeong, Néda, and Barabási [72] have pointed out that the empirically found attachment kernels give rise to theoretical distributions that are inconsistent with empirically observed ones. This finding motivates further research into constructive models of complex networks, aiming to find models that lead naturally to macroscopic properties that are found in real-world applications, while using a simplistic set of microscopic rules that govern their assembly.

One of the limitations of the BA model is the implicit assumption that all newly arriving nodes have perfect information about the topology of the entire network, which may not be the case in real-world applications. To avoid this limitation, a new class of models has been introduced, sometimes known as hidden variable or varying fitness models

[1, 16, 24, 47, 111, 130]. These models are driven by a non-topological quantity, called fitness. The motivation can possibly be best understood from an example. Consider for instance commercial relationships, such as trade or lending. An entrepreneur is unlikely to know all existing relationships of a prospective partner. However, information about non-topological quantities, such as earnings, reputation or market capitalisation are readily available. The hidden variable in this model serves as a proxy for exactly these quantities. Therefore, perfect knowledge is assumed in this class as well. However, knowledge about non-topological quantities is easier to obtain than information about connectivity inside the network.

Caldarelli et al. [24] introduce the concept of a static network model that is driven by node intrinsic fitness. In their model, each node in the network is endowed with a fitness value x, drawn from a probability distribution $\rho(x)$. This fitness is a proxy for the nodes' attractiveness. Edges between pairs of nodes are created with a probability proportional to the attachment kernel $f(x, y)$. It has been shown later that it is always possible to find functions $\rho(x)$ and $f(x, y)$, such that the resulting degree distribution is scale-free [111].

The central variable in [111] is the fitness conditional expectation of the node degree from which the node degree distribution can be derived. In this chapter, another method is illustrated and it will be demonstrated that it is possible to obtain even more detailed expressions of the underlying network and give not just summary statistics, but the exact distribution of the node degree for a node with fitness x. The quantity of central interest here is $p(k, t|x, \tau)$. That is the node degree distribution at time t for a node with fitness x that joined at time $\tau \leq t$. This expression contains all necessary information to fully quantify the topology in this dynamic network model.

Growth models that incorporate fitness as a driver have been studied before [16, 47, 129, 130]. Some of these models are based on a mixture of node degree and fitness to shape the topology of the network. These models are therefore still limited by the argument of perfect information on connectivity that has been developed earlier in this section. Xu et al. [129] and Xu and Zhang [130] illustrate with a mixture of analytical and numerical techniques, that broad degree distributions are tied to a broad fitness distribution

coupled with a properly chosen attachment kernel.

Two problems related to fitness driven networks are the inverse and the direct problem [16, 111], that is to find conditions on $f(x, y)$ or $\rho(x)$ with the respectively other one held fixed, such that the resulting degree distribution is a power-law. Solutions for these two problems in the context of a growing network are presented later in this chapter. Furthermore, it is also pointed out that the distinction of $f(x, y)$ and $\rho(x)$ is over-specifying the problem. It will be illustrated that it is possible to reduce the complexity of the problem to just one quantity, that is the attachment propensity $\lambda(x)$.

This chapter is organised as follows: in the coming section the model is introduced along with the solution of the conditional degree distribution $p(k, t | x, \tau)$. In Section 3, the full analytic theory for this model is presented, including conditions on $f(x, y)$ and $\rho(x)$ to yield arbitrary degree distributions. After that the argument on the redundancy of the classical fitness model is presented and a more compact formulation is introduced. In Section 4, the case of power-law degree distributions is discussed. In Section 5, the calculation of correlation measures of the growing network are presented. Section 6 closes the chapter with a few concluding remarks.

2 The Model

The network under investigation is growing and undirected. The direction of edges is not taken into account because for many real-world networks it is almost impossible to find out the exact joining time order of nodes and the resulting directions of links, once the network has reached a certain age.

During the life-time of the network, in every time step one of two things can happen:

(i) With probability q, a new node is created and endowed with a fitness x drawn from a probability distribution $\rho(x)$ and then joined to the network by adding a link between this new node and one node inside the network. The target node i is chosen with probability proportional to the attachment kernel $f(x, x_i)$, where x_i is

the fitness of the target node.

(ii) With probability $1-q$, a new edge is added between two nodes in the network with probability proportional to $f(x,y)$ where x and y are the fitnesses of the adjacent nodes of the new edge.

The evolution of the network is described in terms of a derived quantity, that normalises $f(x,y)$. That is the probability that a node with fitness x receives a new edge with an adjacent node with fitness y at time t, given by:

$$\gamma(x,y,t) = q \frac{f(x,y)\rho(y)}{\sum_{j=1}^{N(t)} f(x_j,y)} + (1-q)\frac{f(x,y)\sigma(y,t)}{\sum_{i<j}^{N(t)} f(x_i,x_j)}, \tag{3.1}$$

where $\sigma(x,t) = \sum_{i=1}^{N(t)} \delta(x-x_i)$ is the instantaneous fitness density inside the network. The first term accounts for node addition, and the second term for edge addition. In the $t \to \infty$ limit, it holds that $N(t) = qt$ because of the growth rules and therefore also $\sigma(x,t) \to qt\rho(x)$. Consequently, sums can be represented as integrals

$$\sum_{j=1}^{N(t)} f(x,x_j) = \int f(x,\xi)\sigma(\xi,t)d\xi \to qt \int f(x,\xi)\rho(\xi)d\xi, \text{ and} \tag{3.2}$$

$$\sum_{i<j}^{N(t)} f(x_i,x_j) \to \frac{q^2 t^2}{2} \int f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta. \tag{3.3}$$

The limits of integration are the boundaries of the fitness domain. It is therefore a matter of convention. In the following, the unit interval $[0,1]$ will be the assumed range of fitness. If fitness is defined over a different range in a particular application, then it can always be normalised to this interval. Using the continuous limits from above, $\gamma(x,y,t)$ can be rewritten to

$$\gamma(x,y,t) = \frac{1}{t}\left[\frac{f(x,y)\rho(y)}{\int_0^1 f(\xi,y)\rho(\xi)d\xi} + \frac{2(1-q)}{q}\frac{f(x,y)\rho(y)}{\int_0^1\int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta}\right]. \tag{3.4}$$

From the definition of $\gamma(x,y,t)$, one can directly obtain the probability that a node with fitness x will increase its degree by a node addition. This attachment propensity is de-

fined as

$$\lambda(x,t) = \int_0^1 \gamma(x,y,t)dy \tag{3.5}$$

$$= \frac{1}{t}\left[\int_0^1 \frac{f(x,y)\rho(y)}{\int_0^1 f(\xi,y)\rho(\xi)d\xi}dy + \frac{2(1-q)}{q}\int_0^1 \frac{f(x,y)\rho(y)dy}{\int_0^1\int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta}\right]. \tag{3.6}$$

Both of the definitions contain a factor $1/t$, therefore also

$$\lambda(x) = t\lambda(x,t) \tag{3.7}$$

$$\text{and} \quad \gamma(x,y) = t\gamma(x,y,t) \tag{3.8}$$

will be used where applicable. The quantity $\lambda(x)$ will control the topology of the network and integrating over all x leads to the following normalisation condition

$$\int_0^1 \lambda(x)\rho(x)dx = 1 + 2(1-q)/q. \tag{3.9}$$

The factor 2 on the right-hand side comes from the double counting when an edge is added to a pair of nodes, so that two nodes will increase their degree by one.

Using $\lambda(x)$, the quantity of central interest in this chapter $p(k,t|x,\tau)$ can be obtained. $p(k,t|x,\tau)$ is the probability for a node that is created at time τ with fitness x to have a degree k at time t. The governing rate equation in discrete time is given by

$$p(k,t+1|x,\tau) = p(k-1,t|x,\tau)\lambda(x,t) + p(k,t|x,\tau)[1-\lambda(x,t)]. \tag{3.10}$$

The first term on the right-hand side accounts for an increase of the node degree by an edge addition. The second term accounts for an addition to a node with a fitness different from x. The solution of this recurrence relation in two variables k,t can be found using the generating function approach [127]. Define therefore the generating function

$$\tilde{G}(s,t|x,\tau) = \sum_{k\geq 1} s^k p(k,t|x,\tau). \tag{3.11}$$

In the $t \to \infty$ limit, the difference between continuous and discrete time can be ne-

glected, so that Eq. (3.10) can be written in terms of $\tilde{G}(s,t|x,\tau)$ as

$$\frac{\partial \tilde{G}(s,t|x,\tau)}{\partial t} = \lambda(x,t)$$
$$\times \left(\sum_{k \geq 1} s^k p(k-1,t|x,\tau) - \sum_{k \geq 1} s^k p(k,t|x,\tau) \right) \tag{3.12}$$

which is equivalent to the following differential equation

$$\frac{\frac{\partial \tilde{G}(s,t|x,\tau)}{\partial t}}{\tilde{G}(s,t|x,\tau)} = \lambda(x,t)(s-1). \tag{3.13}$$

The solution is given by

$$\tilde{G}(s,t|x,\tau) = s \cdot \left(\frac{t}{\tau} \right)^{(s-1)\lambda(x)}, \tag{3.14}$$

which uses the boundary condition

$$\lim_{t \to \tau+} p(k,t|x,\tau) = \delta_{k1}. \tag{3.15}$$

This boundary condition controls for the fact that a node enters with one single link. Using this boundary gives $\lim_{t \to \tau} \tilde{G}(s,t|x,\tau) = s$ and leads to the solution in Eq. (3.14).

$\tilde{G}(s,t|x,\tau)$ is the generating function for $p(k,t|x,\tau)$. The fitness conditional degree distribution $p(k|x)$ can be obtained by averaging over all entry times $\tau < t$

$$\sum_{k \geq 1} s^k p(k|x) = \frac{s}{(1-s)\lambda(x)+1}. \tag{3.16}$$

Curiously, this expression is time independent and represents therefore a steady state solution to the evolution of the degree distribution. Expanding around $s = 0$ leads to

$$p(k|x) = \frac{1}{1+\lambda(x)} \left(\frac{\lambda(x)}{1+\lambda(x)} \right)^{k-1}. \tag{3.17}$$

Also the average degree of a node with fitness x is easily obtained:

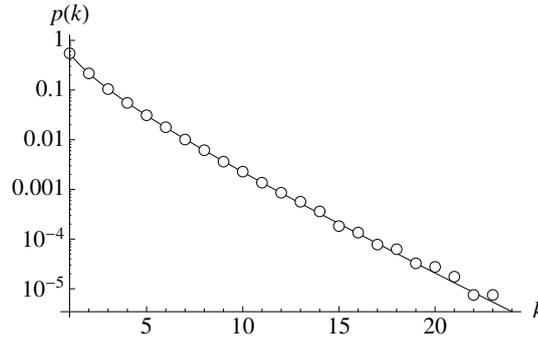$$\bar{k}(x) = 1 + \lambda(x). \tag{3.18}$$

Figure 3.1: Comparison of the analytical result in Eq. (3.20) to results from numerical simulation. The circles mark results from simulation, the solid line the analytical prediction of $p(k)$. Numerical results are obtained from a network simulation over 20,000 time steps, averaged over 20 independent network realisations.

The unconditional degree distribution can then be expressed in terms of $\lambda(x)$ and $\rho(x)$

$$p(k) = \int_0^1 \frac{\rho(x)}{\lambda(x)+1} \left(\frac{\lambda(x)}{1+\lambda(x)} \right)^{k-1} dx. \tag{3.19}$$

Consider now a simple example. For $q = 1$, $f(x,y) = xy$ and standard uniform distributed fitness, $\lambda(x)$ becomes $\lambda(x) = 2x$. It is easily verified that the normalisation condition in Eq. (3.9) is fulfilled. The degree distribution can be expressed in terms of the hypergeometric function $_2F_1(a,b;c;z)$:

$$p(k) = \frac{2^{k-1}}{k} {_2F_1}(k,k,k+1,-2) \tag{3.20}$$

Fig. 3.1 shows a comparison of this result with results from computer simulation of the network growth process. The results are in good agreement.

3　The Design Problem

In the previous section, the necessary theory to derive $p(k)$ for given attachment propensity $\lambda(x)$ and fitness distribution $\rho(x)$ was presented. In this section, the inverse, or design problem will be addressed. That is to find functions $\lambda(x)$ and $\rho(x)$ to yield a given degree distribution $p(k)$. This problem becomes interesting when a practitioner is faced with a network topology and attempts to uncover the underlying governing principles of

the network.

So far, the two functions $\lambda(x)$ and $\rho(x)$ have been distinguished. However, this is not nec-
essary. It becomes, especially in the light of the design problem, clear that the problem
is over-specified with two functions. Therefore, instead of assuming that fitness is some
necessarily quantitative measure, it will only be assumed that fitness can be compared
or ranked. The only axiom that this ranking technology must fulfil is transitivity. In other
words, the ranking must be without contradictions and loops. Instead of the label x for
fitness, each node will then have a rank $u(x)$

$$u(x) = \int^x \rho(\xi)d\xi \qquad (3.21)$$

with $u(x_{\min}) = 0$ and $u(x_{\max}) = 1$. Then $u(x)$ is simply the cumulative distribution func-
tion of x. Since $u(x)$ is monotonically increasing, also the inverse $x(u)$ exists. Using now
this new formalism, the central results from the previous section can easily be rewritten.
The attachment kernel changes to

$$\tilde{f}(u,v)du\,dv = f(x(u), y(v))\rho(x(u))\rho(y(v))dx\,dy. \qquad (3.22)$$

Consequently, $\lambda(x)$ can be rewritten to

$$\lambda(u) = \int_0^1 \frac{\tilde{f}(v,u)}{\int_0^1 \tilde{f}(v,w)dw}dv + \frac{2(1-q)}{q}\frac{\int_0^1 \tilde{f}(v,u)dv}{\int_0^1\int_0^1 \tilde{f}(w,w')dw\,dw'}. \qquad (3.23)$$

The degree distribution in terms of ranking u is then given by

$$p(k) = \int_0^1 \frac{du}{1+\lambda(u)}\left(\frac{\lambda(u)}{1+\lambda(u)}\right)^{k-1}. \qquad (3.24)$$

Another possibility is to label nodes not with their ranking, but with their linking propen-
sity λ. For that, define the effective density $\tilde{\rho}(\lambda) = 1/(d\lambda/du)$. Then, the degree distri-
bution is defined as

$$p(k) = \int_0^\infty \frac{\tilde{\rho}(\lambda)d\lambda}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^{k-1}. \qquad (3.25)$$

This labelling is also closer to the notion of classical network growth models, such as

preferential attachment, where nodes are labelled with their degree which is at the same time the attachment propensity. The connection between the notation in the previous section, which is based on numerical fitness, and the propensity labelling outlined here, can be simply expressed as

$$\tilde{\rho}(\lambda) = \frac{\rho(x)}{\lambda'(x)}\bigg|_{x=\lambda^{-1}(\lambda)}. \tag{3.26}$$

The solution to the design problem is defined as a condition on $\tilde{\rho}(\lambda)$ for given $p(k)$. Define $f(s)$ as the exponential generating function of the degree distribution and substitute the results about the degree distribution from above to get

$$f(s) = \sum_{k \geq 1} \frac{s^k}{k!} p(k) \tag{3.27}$$

$$= \sum_{m \geq 0} \frac{s^{m+1}}{(m+1)!} \int_0^\infty \frac{\tilde{\rho}(\lambda)}{1+\lambda} \left(\frac{\lambda}{1+\lambda}\right)^m d\lambda. \tag{3.28}$$

Notice now, that

$$\sum_{m \geq 0} \frac{s^{m+1}}{(m+1)!} \left(\frac{\lambda}{1+\lambda}\right)^m = \frac{1+\lambda}{\lambda}\left(e^{\frac{s\lambda}{1+\lambda}} - 1\right). \tag{3.29}$$

Then the equation of $f(s)$ can be rewritten to obtain

$$f(s) = \int_0^\infty \frac{\tilde{\rho}(\lambda)d\lambda}{\lambda}\left(e^{\frac{s\lambda}{1+\lambda}} - 1\right). \tag{3.30}$$

Taking first derivative leads to

$$f'(s) = \int_0^\infty \frac{\tilde{\rho}(\lambda)d\lambda}{1+\lambda} e^{s\lambda/(1+\lambda)}. \tag{3.31}$$

Using an integral representation of the Dirac-delta function

$$\delta\left(\eta - \frac{\lambda}{1+\lambda}\right) = \int_{-\infty}^\infty \frac{d\sigma}{2\pi} e^{-i\sigma\eta + i\sigma\frac{\lambda}{1+\lambda}}, \tag{3.32}$$

and that $\int f(x)\delta(g(x))dx = f(x_0)/|g'(x_0)|$ leads to the solution

$$\tilde{\rho}(\lambda) = \frac{1}{1+\lambda} \int_{-\infty}^\infty \frac{d\sigma}{2\pi} f'(i\sigma)e^{i\sigma\frac{\lambda}{1+\lambda}}. \tag{3.33}$$

This equation specifies the necessary form of $\tilde{\rho}(\lambda)$ to induce an arbitrary degree distribu-

tion $p(k)$. The corresponding conditions on $\lambda(x)$ and $\rho(x)$ can be found by substituting the definition of $\tilde{\rho}(\lambda)$ in Eq. (3.26).

In the following section a use case of this formalism will be illustrated, and briefly compared to a direct approach that is not making use of the general solution shown here.

4   Scale-free Networks

Scale-free networks, those with a power-law degree distribution, can be found in many real-world applications. This has already been highlighted in the introductory chapter. To solve for the necessary conditions, three different approaches are discussed in the following section. One is utilising the technology that is derived in the previous section. The two others are direct approaches that are tailored specifically to the power-law distribution.

In order to find the condition on $\tilde{\rho}(\lambda)$ using Eq. (3.33), the exact form of $f(s)$ for a power-law needs to be found. That is

$$f(s) = \sum_{k \geq 1} \frac{As^k}{k^\alpha k!} \tag{3.34}$$

Where A is a normalisation constant. The corresponding integral representation is given by

$$f(s) = A \int_0^\infty t^{\alpha-1} \left(e^{se^{-t}} - 1 \right) dt. \tag{3.35}$$

After obtaining $f'(s)$ using Eq. (3.33) for $\tilde{\rho}(\lambda)$, and substituting the normalisation condition $\int_0^\infty \tilde{\rho}(\lambda)d\lambda = 1$, the condition for a power-law degree distribution on $\tilde{\rho}(\lambda)$ can be obtained:

$$\tilde{\rho}(\lambda) = \frac{1}{\Gamma(\alpha)\zeta(\alpha)} \cdot \frac{1}{1+\lambda} \ln\left(\frac{1+\lambda}{\lambda}\right)^{\alpha-1}. \tag{3.36}$$

The corresponding forms of $\lambda(x)$ and $\rho(x)$ can be found by substituting the definition of $\tilde{\rho}(\lambda)$ into Eq. (3.36).

But there exists also another way to directly obtain conditions on $\rho(x)$ and $\lambda(x)$ so that the resulting degree distribution is scale-free. As a first step, average the generating func-

tion in Eq. (3.16) over all fitness and factor out $1/\lambda(x)$, to obtain

$$\sum_{k \geq 1} s^k \int_0^1 p(k|x)\rho(x)dx = \int_0^1 \frac{\rho(x)}{\lambda(x)} \frac{s}{\frac{1+\lambda(x)}{\lambda(x)} - s} dx. \tag{3.37}$$

The generating function of a power-law distribution and its integral representation are given by

$$\sum_{k \geq 1} As^k k^{-\alpha} = A\frac{s}{\Gamma(\alpha)} \int_0^\infty \frac{t^{\alpha-1}}{e^t - s} dt \tag{3.38}$$

The right-hand side is the integral representation of the polylogarithm function. The condition on $\lambda(x)$ and $\rho(x)$ can therefore be expressed in terms of the corresponding generating functions. It must hold that

$$\int_0^1 \frac{\rho(x)}{\lambda(x)} \frac{s}{\frac{1+\lambda(x)}{\lambda(x)} - s} dx = A\frac{s}{\Gamma(\alpha)} \int_0^\infty \frac{t^{\alpha-1}}{e^t - s} dt. \tag{3.39}$$

The problem of obtaining a power-law degree distribution with the described mechanism is therefore reduced to the solution of this equation. The left-hand side and right-hand side are identical if one substitutes on the left-hand side

$$\frac{1+\lambda(x)}{\lambda(x)} = e^t \quad \text{and} \quad \frac{\rho(x)}{\lambda(x)}dx = -\frac{A}{\Gamma(\alpha)} t^{\alpha-1}dt \tag{3.40}$$

Using these, it is now clear that the problem of solving the integral equation in (3.39) can be transformed into the problem of solving the following differential equation which arises from the two expressions in Eq. (3.40)

$$\rho(x) = \ln\left(\frac{1+\lambda(x)}{\lambda(x)}\right)^{\alpha-1} \cdot \frac{\lambda'(x)}{1+\lambda(x)} \cdot \frac{A}{\Gamma(\alpha)}. \tag{3.41}$$

the parameter A can be found by the normalisation condition $\int_0^1 \rho(x)dx = 1$, since $\rho(x)$ is a density function. That leads to $A = \zeta^{-1}(\alpha)$, whereby $\zeta(x)$ is the standard Riemann-zeta function. This expression can be found by substituting $z = \ln([1+\lambda(x)]/\lambda(x))$ and using the definition of the polylogarithm function

$$\zeta(\alpha) = \text{Li}_\alpha(1) = \frac{1}{\Gamma(\alpha)} \int_0^\infty \frac{t^{\alpha-1}}{e^t - 1} dt \tag{3.42}$$

Eq. (3.41) represents the solution of inverse problem [111], that is to find a function $\rho(x)$ for given $f(x, y)$, such that the degree distribution of the resulting network is a power-law.

The solution to the direct problem [111], that is to find conditions on $f(x, y)$ for given $\rho(x)$ such that the degree distribution follows a power-law, can be found as the solution of the implicit differential equation in Eq. (3.41), which can be obtained by integration:

$$\zeta(\alpha)\Gamma(\alpha)\int^{x}\rho(\xi)d\xi = \int_{0}^{\ln\left(\frac{1+\lambda(x)}{\lambda(x)}\right)}\frac{v^{\alpha-1}}{1-e^{v}}dv \tag{3.43}$$

This equation can be solved numerically for $\lambda(x)$ for arbitrary values of x and given $\rho(x)$. Notice that if one substitutes the normalisation condition on $\lambda(x)$:

$$\int_{0}^{1}\lambda(x)\rho(x)dx = 1 + \frac{2(1-q)}{q} \tag{3.44}$$

into Eq. (3.41), one obtains

$$\frac{2}{q} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}. \tag{3.45}$$

This equation restricts the interval of possible solutions to $\alpha \in [2, 2.47875\ldots]$. For $q = 1$, α is on the righthand boundary of this interval. Interestingly, this is very close to several real-world networks. For example, the out-exponent of the world wide web is 2.45 [4]. The network of internet routers has in- and out-exponent of 2.48 [52]. But this is an artefact that does not only arise from this model, but also from every class of models that lead to an exact power-law: assuming that the degree distribution follows a strict power-law, that is $P(k) = Ck^{-a}$, and that a single node with one adjacent edge is added in each time step leads to an average degree $\bar{k} = 2$. C can be obtained from normalisation: $C \cdot \sum_{k} k^{-a} = 1$ and therefore $C = \frac{1}{\zeta(a)}$. Using now the average degree, one obtains $2 = \frac{1}{\zeta(a)}\sum_{k} k \cdot k^{-a}$ which is equivalent to $2 = \frac{\zeta(a-1)}{\zeta(a)}$ and hence $\alpha = 2.47875\ldots$

One other aspect that can be learned from the calculation above is that the power-law degree distribution can only be obtained if $\lambda(x)$ is singular at $x = 1$. It should be noticed that the calculations above use the assumption that the range of $\lambda(x)$ is $[0, \infty)$. That has been used for the step in Eq. (3.40). To equate both sides of the expression in the pre-

ceeding equation, λ needs to vary from 0 to ∞ when x goes from 0 to 1. Therefore, $\lambda(x)$ needs to have a singularity at $x = 1$, but it must be an integrable singularity, because it has also been shown that $\int_0^1 \lambda(x)\rho(x)dx$ is finite.

To confirm the analytical result, numerical simulations have been carried out for both, the solution of the inverse and the direct problem. The results can be found in Figs. 3.2 and 3.3.
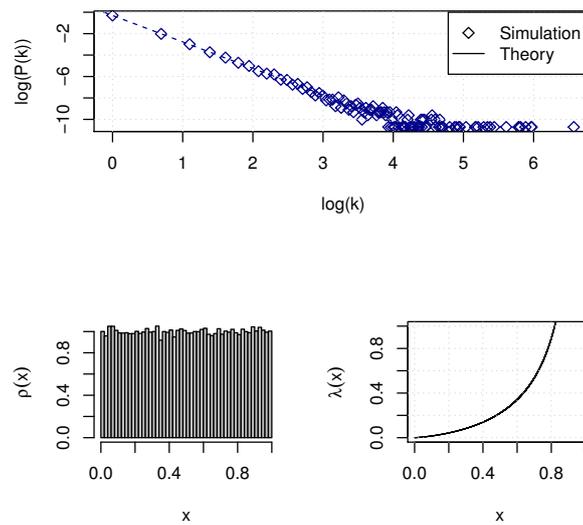
Figure 3.2: Results for flat $\rho(x)$ and $\lambda(x)$ according to the theory. The theoretical power-law fits well to the numerical results.

5   Fitness Correlations

It has been shown that the topological structure of the static fitness network can be obtained by employing fitness-fitness correlations, which are identical to the properly normalised attachment kernel in the case of the static network [24, 111]. The fitness-fitness correlations for the growing network are different but can be obtained from first principles. For brevity, the correlations are derived in terms of ranking rather than fitness. The transition to the fitness based notation can be found by simple substitution. Define the
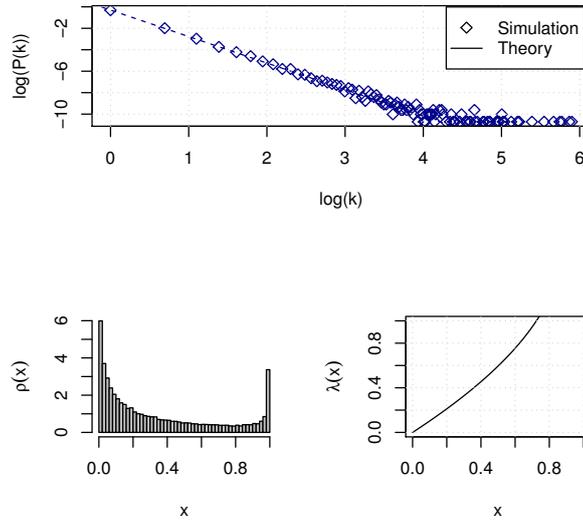
Figure 3.3: Results for $\lambda(x) = x/(1-x)^{0.25}$ and $\rho(x)$ according to the theory. The numerical results for the degree distribution match the theory well.

probability that a link exists between two nodes with rank u and v at time t as

$$C(u,v,t) = \mathbb{E}\left[\frac{1}{2t}\sum_{i,j=1}^{N(t)} C_{ij}(t)\delta(u-u_i)\delta(v-v_j)\right]. \tag{3.46}$$

$\delta(x)$ is the Dirac delta function and $C_{ij}(t)$ is the adjacency matrix of the network at time t. The sum is normalised to the number of edges in the network. Since at each time step either a node with an edge or only an edge is added to the network, the number of edges at time t is just t. Since the sum runs over all pairs of (i,j) and the network is not directed, each edge will be counted twice, once in the upper and once in the lower triangle of the adjacency matrix. To account for this, the factor $1/2$ leads the sum-term. Similar to the methodology in the previous sections, also $C(u,v,t)$ can be found by recursion. It obeys

$$\begin{aligned}
C(u,v,t+1) &= C(u,v,t)\left(1-\frac{1}{t}\right) + \frac{q}{2t}\left(\frac{f(u,v)}{\int f(w,v)dw} + \frac{f(u,v)}{\int f(u,w)dw}\right) \\
&\quad + \frac{1-q}{t}\frac{f(u,v)}{\int\int f(w,x)dw\,dx}.
\end{aligned} \tag{3.47}$$

The first term accounts for the change of the normalisation of $C(u,v,t)$ when a new edge is added. This term arises from the factorisation

$$\frac{1}{t+1} = \frac{1}{t}\frac{1}{1+\frac{1}{t}} = \frac{1}{t}\left(1 - \frac{1}{t} + \frac{1}{t^2} - \frac{1}{t^3} + \cdots\right).$$ (3.48)

Thus, in the $1/t$ approximation, $1 - 1/t$ is left. The second and third term of Eq. (3.47) account for the increase of $C(u,v,t)$ by addition of an edge, either through node or direct edge addition. The normalising factor $1/2$ accounts for the double counting of edges that happens by summing once in row-direction and once in column-direction. Define further

$$D(u,v) = \frac{q}{2}\left(\frac{f(u,v)}{\int f(w,v)dw} + \frac{f(u,v)}{\int f(u,w)dw}\right), \quad \text{and}$$ (3.49)

$$E(u,v) = (1-q)\frac{f(u,v)}{\int\int f(w,x)dw\,dx}.$$ (3.50)

The solution of Eq. (3.47) is then

$$C(u,v,t) = D(u,v) + E(u,v)\frac{t-1}{t}.$$ (3.51)

The importance of the initial condition is thus decaying as $1/t$ and therefore a steady state fitness correlation matrix can be found:

$$C(u,v) = D(u,v) + E(u,v).$$ (3.52)

6 Conclusion

A full theory on the degree distribution of growing networks with static node intrinsic fitness has been derived in this chapter. For any given two of the three central functions $\lambda(x), \rho(x)$ and $p(k)$, the third one can be directly derived. It has already been argued in the introduction that the class of growing fitness induced networks is of special interest. This is because the information requirement for each agent is more realistic than in the degree preferential models for many applications. Previously, these networks have al-

ready been studied for exactly this reason. However, those studies were limited to static networks. Growth is an essential addition to these models, as many structures around us grow naturally in time.

Methodologically, it has been shown that instead of the attachment kernel $f(x, y)$, the attachment propensity $\lambda(x)$ is driving the network evolution, coupled with the fitness density $\rho(x)$. Furthermore, it has been illustrated that the distinction between $\lambda(x)$ and $\rho(x)$ is from a theoretical point of view over-specifying the problem. Just one function is enough, for instance the distribution of attachment propensities $\tilde{\rho}(\lambda)$. However, it is acknowledged that in some applications the distinction between $\rho(x)$ and $\lambda(x)$ is important: not at last in a situation when an understanding of the internal mechanisms is required but only two of the three functions $\lambda(x), \rho(x), p(k)$ are known.

As an example, the particular case of power-law degree distributions has been discussed. It has been shown that networks with power-law degree distributions arise only under very specific circumstances within this set of rules. This will motivate further research in the area, exploring realistic models, that lead naturally to the ubiquitous power-law degree distributions combined with node intrinsic fitness.

Chapter 4

Percolation of Failures on Complex Networks

The topology of growing networks with node intrinsic fitness was discussed in the previous chapter. Within that class of networks, fitness serves as a measure of attractiveness and it has been illustrated how the knowledge of the distribution of fitness and the attachment propensity can be translated into the resulting degree distribution of a growing network. In this chapter, the functionality of networks will be investigated in more detail. Here, fitness does not only serve as a measure of attractiveness, but also as a measure of a node's robustness against failure. The probability that a node fails increases with the number of failures in its direct neighbourhood and decreases with higher fitness. Both static and dynamic network models are considered. Analytical results for the percolation threshold and the occupied fraction are derived. The distinction between static and dynamic networks, different network densities as well as varying attachment kernels is discussed and their effect on percolation investigated.

1 Introduction

The investigation of contagious processes is a vital part of many fields in academia, whether it is the transmission of sexual diseases, studied in social science [81], the spread of financial distress in economics [15, 105] or the spread of viruses in epidemiology [27]. The most simple transmission models assume perfect mixing of the underlying population. In other words, contacts between banks, sexual partners or pedestrians are purely random. This simplification is made in order to be able to formulate a system of coupled nonlinear differential equations that can be solved with ease [91]. However, this assumption is an oversimplification of the underlying problem. Contact patterns are largely heterogeneous [3]. This heterogeneity has been endogenised by considering underlying networks that describe different contact motifs [78, 86, 97, 110].

Beyond the introduction of a topology of contacts, further aspects have been introduced, such as timing and local transmission probabilities [91], in order to account for real-life phenomena. The spreading behaviour in multi-layered network architectures and networks of networks has also been investigated [22, 41, 69, 132]. Other features, such as awareness and vaccinations were considered too [61, 98]. In other studies, more complex models that attempt to explain default cascades, occurring for instance in financial markets, have been considered [105]. Roukny et al. [105] set up a model that takes different kinds of shocks, as well as correlations, not only degree-degree but also degree-robustness, and a model specific market illiquidity parameter into account.

In this chapter, a flat network is considered. Flat in the sense that it consists of one layer. The underlying topology is not chosen to be arbitrary, it is derived from a hidden variable model [21, 24, 111]. In contrast to node degree based models, such as the Barabási-Albert model [8], or models, that combine local attractiveness together with node degree as an attractor for new edges [19], fitness/hidden-variable models are purely driven by static node intrinsic fitness. The two functions that determine the topology of the network and all related quantities are the attachment kernel $f(x, y)$, that describes the probability that a node with fitness x originates an edge towards a node with fitness y, and the proba-

bility density $\rho(x)$, that describes the distribution of fitness in the system. The literature on pure fitness models is split into two parts, static and dynamic models, which are both considered in this chapter. Static networks have been identified as the correct model in various applications such as the interbank lending market and the world trade network [37, 57, 58]. Their dynamic counterpart, that is discussed in the previous chapter, relies on assumptions that are met in several real-world examples. As it has already been argued in the preceding chapter, the attractiveness of fitness driven models is the amount of information that a local agent is assumed to have. While degree dependent growth models assume that new nodes have information about the connectivity of the entire network, it is sufficient for fitness dependent models to assume that nodes have information about the ranking of some derived, topological independent quantity. Consider for instance the network of investments. The balance sheet structure of an enterprise is more likely to be accessible to an investor than the absolute number of other investors invested in a particular company.

In this chapter, attractiveness and resilience are directly coupled. Fitness determines the propensity to acquire more edges in the network, but also measures robustness against failures in the direct neighbourhood of a node. This association of attractiveness and robustness is not just interesting as an exercise for a theoretician, it is of great interest to organizations that can influence the underlying free parameters, such as the distribution of fitness, as well as the average node-degree in the system. Consider for example the interbank lending market. Assuming that the selection process is fixed, in the sense that $f(x, y)$ is given and cannot be influenced, the policy maker or regulator can create incentives to stimulate interaction between existing agents in the form of new edges that increase the average node degree, or they can reshape the fitness distribution $\rho(x)$ by the introduction of a tax. It is shown later in this chapter that these two quantities largely influence the percolation threshold as well as the fraction of the network that is occupied by the giant vulnerable component.

Attractiveness and robustness cannot always be assumed to correspond one-to-one, such as in the case of the sexual contact network that has been investigated in [81]. However there are many scenarios in which this correspondence appears realistic, consider again

the interbank lending market. Here, attractiveness is derived directly from the likelihood to file for bankruptcy. Another possible model is the network of buyer-supplier relationships in the technological sector. Think of the arrival of a new technology as an epidemic process. In this case large suppliers are attractive due to their size, but also control the adoption of new technologies. One example is the Flash-technology, that has never been adopted by Apple on its mobile devices.

The remainder of this chapter is organised as follows: in Section 2 the two models under investigation are presented. In Section 3 the core results on percolation for the given models are derived. In Section 4 the main results are presented. Section 5 closes the chapter with concluding remarks.

2 Models

In this chapter the exposure of contact networks to random failures is studied. Two different classes of network formation models are considered. These are static fitness models [24] and their dynamic counterpart that was discussed in the previous chapter. The dynamic model is characterised by a constant addition of nodes, while the static model comprises a fixed number of nodes. The timescale of the network formation is assumed to be longer than the typical timescale of the spread of a failure, such that the formation of the network and its occupation with a large failed component can be considered separately. Both variations of the fitness model are characterised by an attachment kernel $f(x, y)$ and a fitness distribution $\rho(x)$. The attachment kernel $f(x, y)$ describes the probability that a node with fitness x originates a new edge toward a node with fitness y. The fitness of each node is static over the lifetime of the network and is drawn from the probability density $\rho(x)$. Fitness is assumed to be distributed over the unit interval $[0, 1]$. If fitness is distributed over another interval in a particular application it can always be normalised to the interval $[0, 1]$.

The networks under consideration are undirected and every agent is in one of two states, bankrupt or solvent / dead or alive / technology-adapter or -refuser, / infected or sus-

ceptible etc. Although the model can be understood in terms of many applications, it is in general a binary rule as it was introduced in [124]. The two states will be referred to as solvent and bankrupt hereafter. Every agent i is initially solvent and changes its state to bankrupt if a critical fraction of the agents in the agent's neighbourhood have gone bankrupt. This critical fraction is given by its fitness x_i. If at least a fraction x_i of a node's neighbours has changed its state to bankrupt, node i also changes its state from solvent to bankrupt. In the context of the interbank lending market, an edge between two banks represents a shared exposure. Commercial relationships are assumed to be two-sided so that a bankruptcy on one end of the edge will always affect the node on the other end. If too many banks that have entered a commercial relationship with bank i fail, also bank i will fail. The critical fraction is hence coupled to the degree of a node. Consider now a randomly induced initial bankruptcy somewhere in the network. This bankruptcy can only propagate to a neighbour that has a fitness value such that $x_i \leq 1/k_i$. Nodes that fulfil this condition are referred to as vulnerable. Once the initially solvent network is perturbed with the state change of a single node to bankruptcy, the spread of bankruptcies over the network develops asynchronously in accordance with the simple threshold rule $x_i \leq 1/k_i$.

The vulnerability condition $x \leq 1/k$ implies that a bank distributes its exposure uniformly over all its associated business partners, which might be different in the real world but it will be abstracted from this possibility for the sake of simplicity. If the vulnerability condition is fulfilled, a single failing bank in the neighbourhood of a node i can cause i to fail as well.

2.1   Static Model

The static fitness model describes a network that comprises a fixed number of N nodes and M edges. The quantity of central interest here is the degree of a node. Analytical results on the degree distribution for this model have been obtained earlier, see for instance [24, 111]. In order to consider percolation in this network, it is necessary to obtain the fitness-conditional out-degree distribution. This quantity, also referred to as the

propagator has been found to be Poissonian in [21]. While the propagator is derived constructively by Boguñá and Pastor-Satorras [21], a derivation from first principles is described below.

Although the network is static i.e. it contains a fixed number of nodes, the process in which edges are added can be understood as a sequential procedure. Edges are deployed one by one, not all at the same time. An edge is added to a pair of nodes (i, j) with probability $f(x_i, x_j)/\sum_{k,l} f(x_k, x_l)$. The networks under consideration are sparse, in the sense that $M \ll N^2$, thus the possibility of adding an edge to a pair of nodes that is already connected is negligible in the thermodynamic limit. The probability that a node with fitness x increases its degree by one during an edge-addition step is defined as

$$\lambda(x, N) = \int_0^1 \frac{f(x, y)\sigma(y, t)}{\sum_{i<j}^N f(x_i, x_j)} dy. \qquad (4.1)$$

The same as in the last chapter $\sigma(y, t) = \sum_{i=1}^N \delta(y_i - y)$ is the instantaneous fitness distribution, which is here directly $N\rho(y)$. In the $N \to \infty$ limit, the following integral representation can be obtained

$$\lambda(x, N) = \frac{2}{N} \int_0^1 \frac{f(x, y)\rho(y) dy}{\int_0^\infty \int_0^\infty f(\xi, \eta)\rho(\xi)\rho(\eta) d\xi d\eta}. \qquad (4.2)$$

Since $1/N$ is a factor of $\lambda(x, N)$, also define $\lambda(x) = N\lambda(x, N)$. Using the postulate that the edge addition procedure can be understood as a sequential process, the fitness-conditional degree distribution can be found using a rate equation approach. The probability that a node with fitness x has degree k in a network with N nodes and M edges will be denoted with $p_{M,N}^{(s)}(k|x)$ and evolves as

$$p_{M+1,N}^{(s)}(k|x) = p_{M,N}^{(s)}(k|x)[1 - \lambda(x, N)] + p_{M,N}^{(s)}(k-1|x)\lambda(x, N). \qquad (4.3)$$

The first part of Eq. (4.3) corresponds to the an edge update that occurs at a node with fitness unequal to x, the second part corresponds to an edge update that increases the degree of a node with fitness x by one. The recurrence relation in Eq. (4.3) can be solved

using a generating function approach. Define

$$F_{M,N}^{(s)}(s|x) = \sum_{k \geq 0} p_{M,N}^{(s)}(k|x)s^k. \tag{4.4}$$

Multiplying Eq. (4.3) with s^k and summing over k leads to

$$F_{M+1,N}^{(s)}(s|x) - F_{M,N}^{(s)}(s|x) = F_{M,N}^{(s)}(s|x)\lambda(x,N)(s-1). \tag{4.5}$$

For large M, Eq. (4.5) can be approximated as an ordinary differential equation in M with initial condition $F_{0,N}^{(s)}(s|x) = 1$. The initial condition arises from the observation that in a network without edges, the conditional degree distribution is peaked at zero: $p_{0,N}^{(s)}(k|x) = \delta_{k0}$, where δ_{xy} is the Kronecker delta. Using this, and that $1/N$ is a factor of $\lambda(x,N)$, leads to

$$F_M(s|x) = e^{M/N \lambda(x)(s-1)}. \tag{4.6}$$

Expanding Eq. (4.6) in s around $s = 0$ leads to the expression for the fitness dependent degree distribution

$$p_{M,N}^{(s)}(k|x) = \frac{e^{-M/N \lambda(x)}(\frac{M}{N}\lambda(x))^k}{\Gamma(k+1)}. \tag{4.7}$$

This result is sufficient to characterize vulnerable nodes in the network. The probability that a randomly chosen node has degree k and is vulnerable, i.e. whose fitness value is less than the reciprocal of its degree, is defined as $Q^{(s)}(k) = \mathbb{P}[k_i = k \cap x_i \leq 1/k]$, where $\mathbb{P}[A]$ is the probability of an event A. $Q^{(s)}(k)$ can be expressed in terms of $p_{M,N}^{(s)}(k|x)$:

$$Q^{(s)}(k) = \int_0^{1/k} p_{M,N}^{(s)}(k|x)\rho(x)dx \tag{4.8}$$

$$= \frac{1}{\Gamma(k+1)} \int_0^{1/k} e^{-M/N \lambda(x)} \left(\frac{M}{N}\lambda(x)\right)^k \rho(x)dx. \tag{4.9}$$

This expression is in excellent agreement with results from numerical simulations of the network assembly process. Fig. 4.1 compares the prediction of Eq. (4.9) with numerical simulations for various cases of $\rho(x), f(x,y)$ and $\langle k \rangle$. Another observation from Fig. 4.1 is that the qualitative differences between $Q^{(s)}(k)$ for exponential- and Pareto distributed fitness are small. The qualitative differences between random attachment, i.e. $f(x,y) = 1$
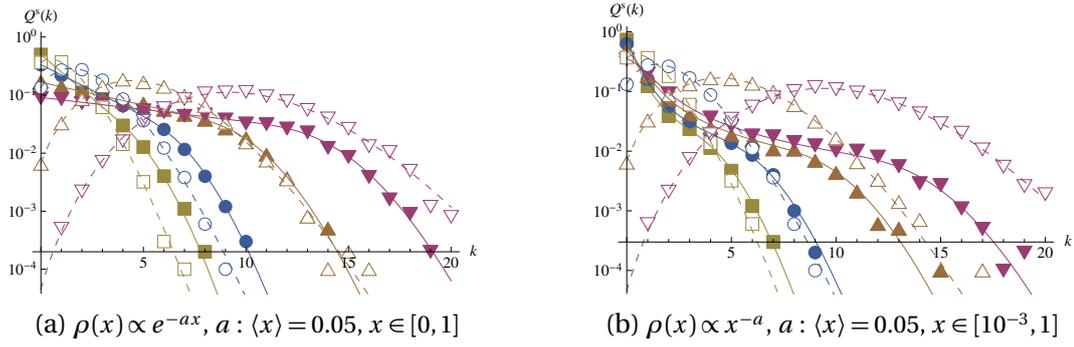
(a) $\rho(x) \propto e^{-ax}$, $a : \langle x \rangle = 0.05$, $x \in [0,1]$ (b) $\rho(x) \propto x^{-a}$, $a : \langle x \rangle = 0.05$, $x \in [10^{-3}, 1]$

Figure 4.1: The plots show $Q^{(s)}(k)$ for different fitness distributions: (a) shows different configurations with fitness distributed exponentially on the unit-interval, with average fitness $\langle x \rangle = 0.05$. (b) shows different configurations with Pareto distributed fitness on the interval $[10^{-3}, 1]$, whereby the zero is excluded to avoid the singularity. Random selection, i.e $f(x,y) = 1$ is marked with hollow symbols, mutual attractiveness, i.e. $f(x,y) = xy$ with solid markers. The different plot markers, distinguish different average degrees: $\langle k \rangle = 1$ squares (yellow), $\langle k \rangle = 2$ circles (blue), $\langle k \rangle = 5$ triangles (brown), $\langle k \rangle = 10$ downward triangles (purple). The lines indicate analytical results from Eq. (4.9), whereby solid lines represent the case of $f(x,y) = xy$ and dashed lines $f(x,y) = 1$. Numerical results are obtained in a network with $N = 10^4$ nodes.

and mutual attractiveness with $f(x,y) = xy$ become more evident for large $\langle k \rangle$. The average degree changes the behaviour of $Q^{(s)}(k)$ significantly. The distribution becomes generally broader for higher $\langle k \rangle$. One more interesting aspect of the network is the effect of diversification. Fig. 4.2a shows the conditional probability that randomly chosen node is vulnerable, given that it has degree k. That is

$$\mathbb{P}[x \leq 1/k \,|\, k_i^{\text{out}} = k] = \frac{Q(k)}{P(k)} \tag{4.10}$$

with $P(k) = \int_0^1 p_{M,N}(k|x)\rho(x)dx$. Fig. 4.2a shows that for the case of mutual attractiveness, the transition between finding a vulnerable node with certainty and not finding any vulnerable node for a given degree is occurring at higher degrees for larger $\langle k \rangle$. This implies that adding new edges destabilizes the system. This adverse diversification effect is at first glance counterintuitive. It implies that the higher the average degree is in the system, the larger is the probability that a node with a given degree is vulnerable. In other words, given a node with degree $k' > 1$, it is more likely that this node is vulnerable in a system with $\langle k \rangle = k'$, than in a system with $\langle k \rangle = 1$. This is because in a system with average degree $\langle k \rangle = 1$, a node has to have a greater than average fitness to attract k' edges, and therefore is more unlikely to fail. In a system with $\langle k \rangle = k'$ however, a node only needs to have average fitness to attract k' many edges. Due to its lower fitness, it is therefore more prone to failures. The case of random attachment, i.e. $f(x,y) = 1$ shows
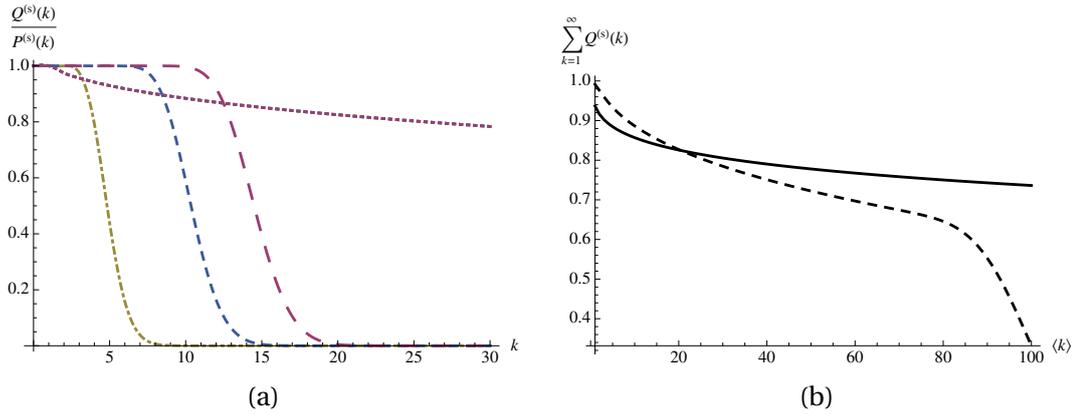
Figure 4.2: Plot (a) shows degree-conditional probabilities of vulnerability for different average degrees $\langle k \rangle = 1$: dot-dashed (mustard), $\langle k \rangle = 5$: fine-dashed (blue), $\langle k \rangle = 10$ coarse-dashed (purple). Examples for $f(x,y) = xy$ are indicated with different dashing patterns, random attachment $f(x,y) = 1$ as dotted lines. The dotted lines are all overlapping, indicating that for random attachment, the conditional probability to find a vulnerable node is independent of $\langle k \rangle$. This is analytically confirmed in Eq. (4.11). Plot (b) illustrates the relationship between the fraction of vulnerable nodes in the network and the average degree $\langle k \rangle$. The configuration in this example is power-law fitness, i.e. $\rho(x) = x^{-a}$, $x \in [10^{-3}, 1]$ $a : \langle x \rangle = 0.05$. The solid line shows the result for $f(x,y) = xy$, the dashed line represents random attachment, i.e. $f(x,y) = 1$.

an entirely different behaviour. The quotient $Q^{(s)}(k)/P^{(s)}(k)$ is the same over the whole range of k. This arises from the form of $\rho(x)$, for $f(x,y) = 1$, the quotient is given by

$$\frac{Q(k)}{P(k)} = \int_0^{1/k} \rho(x)dx. \tag{4.11}$$

The case of random attachment is different because fitness does not have an influence on the degree of a node. The probability for a randomly chosen node to be vulnerable simply decays therefore with the cumulative distribution of fitness as shown in Eq. (4.11). Fig. 4.2b shows that adding additional edges to the network has a positive effect overall, as expected. The fraction of vulnerable nodes decays with increasing $\langle k \rangle$.

Other quantities of interest are the conditional and the unconditional expectation of the degree of a randomly chosen node $\mathbb{E}[k|x]$ and $\langle k \rangle$ respectively. These are obtained as

$$\mathbb{E}[k|x] = \sum_{k \geq 0} k p_{M,N}^{(s)}(k|x) = \frac{M}{N} \lambda(x) \quad \text{and} \tag{4.12}$$

$$\langle k \rangle = \int_0^1 \mathbb{E}[k|x]\rho(x)dx = \frac{M}{N} \int_0^1 \lambda(x)\rho(x)dx. \tag{4.13}$$

Since the out-degree of a node is determined by twice the number of edges M divided by the number of nodes that are in the network N, Eq. (4.13) implies a normalization

condition for $\rho(x)$ and $\lambda(x)$, such that

$$\int_0^1 \lambda(x)\rho(x)dx = 2. \tag{4.14}$$

Therefore, it holds that $\langle k \rangle = 2M/N$.

2.2 Dynamic Model

The dynamic model, as opposed to the static model, is characterised by sequential additions of nodes and edges. Thus, the discrete time, Markov chain approach that is used in the previous subsection is employed here again. The dynamic model that is investigated here is the same as the one in the previous chapter. Therefore only the most central parts of the derivation will be repeated here. The reader is referred to Chapter 3 for more details.

The dynamic network model is characterised by the following dynamic rule. At each time step, one of two things can occur:

(i) With probability q, a new node is created and endowed with a fitness value x, drawn from a probability density $\rho(x)$. A node inside the network is then chosen randomly to form a new edge with this new node. The target of this edge depends mutually on the new node's fitness x and the target node's fitness y. The probability for such an edge is proportional to $f(x,y)$.

(ii) With probability $1-q$, a new edge between two existing nodes inside the network is created. The probability for a such an edge depends again on the node's fitness and is proportional to $f(x,y)$.

The central quantity of interest is the probability that a node which joined the network at time τ with fitness x has degree k at time t. This probability will be expressed with $p_t^{(d)}(k|x,\tau)$. The evolution of this quantity can be expressed in terms of the attachment propensity $\theta_q(x)$, that has been elaborated in the previous chapter. Define the probabil-

ity that a node with fitness x receives a node edge at time t as

$$\theta_q(x,t) = \frac{1}{t}\left[\int_0^1 \frac{f(x,y)\rho(y)}{\int_0^1 f(\xi,y)\rho(\xi)d\xi}dy + \frac{2(1-q)}{q}\frac{\int_0^1 f(x,y)\rho(y)dy}{\int_0^1\int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta}\right]. \quad (4.15)$$

The first part of corresponds to a node addition step. The term inside the integral is
the properly normalised probability that the node with fitness x initiates an edge toward
the new node, averaged over all possible fitness of the newly added node. The second
term corresponds to the case of edge addition. The fraction is the properly normalised
probability that a node with fitness x receives a new edge. The sums run over all nodes
that are present in the network at time t. Also here, only the case of a sparse network is
considered, such that the number of edges is at all times $O(N(t))$. Thus, the problem of
edge duplications can be neglected in the leading order approximation. Since $1/t$ is a
factor of $\theta_q(x,t)$ in the $t \to \infty$ limit, also $\theta_q(x) = t\theta_q(x,t)$ will be used in the following.
The evolution of the out-degree distribution of a node with fitness x that joined at τ,
$p_t^{(\mathrm{d})}(k|x,\tau)$ obeys

$$p_t^{(\mathrm{d})}(k|x,\tau) = p_t^{(\mathrm{d})}(k|x,\tau)(1-\theta_q(x,t)) + p_t^{(\mathrm{d})}(k-1|x,\tau)\theta_q(x,t). \quad (4.16)$$

The first term in Eq. (4.16) corresponds to the situation in which an edge is added to
a node that has a fitness different from x. The second term corresponds to an update
around a node with fitness x. The solution of Eq. (4.16), averages over all entry times
$\tau < t$ is given by

$$p^{(\mathrm{d})}(k|x) = \frac{1}{1+\theta_q(x)}\left(\frac{\theta_q(x)}{1+\theta_q(x)}\right)^{k-1}. \quad (4.17)$$

The probability that a randomly chosen node has degree k and is vulnerable is calculated
in the same way as in the previous subsection and is found as

$$Q^{(\mathrm{d})}(k) = \int_0^{1/k} p^{(\mathrm{d})}(k|x)\rho(x)dx \quad (4.18)$$

$$= \int_0^{1/k}\left(\frac{\theta_q(x)}{1+\theta_q(x)}\right)^{k-1}\frac{\rho(x)dx}{1+\theta_q(x)}. \quad (4.19)$$

(a) $\rho(x) \propto e^{-ax}$, $a : \langle x \rangle = 0.05$, $x \in [0, 1]$
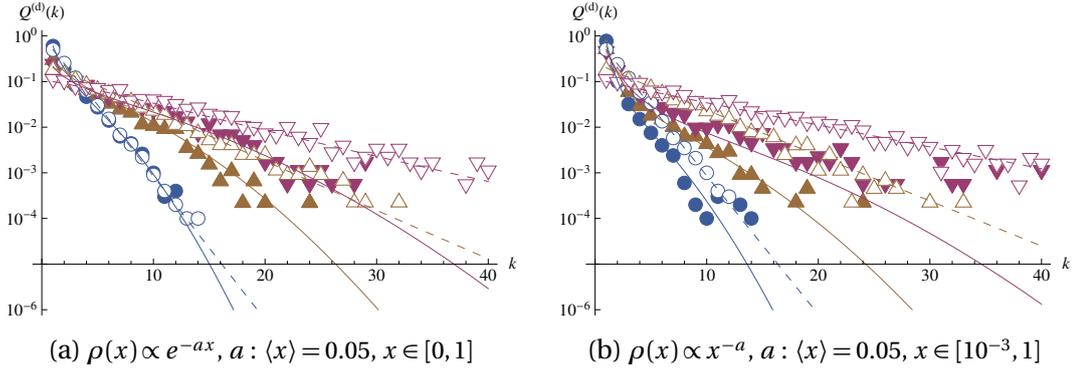(b) $\rho(x) \propto x^{-a}$, $a : \langle x \rangle = 0.05$, $x \in [10^{-3}, 1]$

Figure 4.3: The plots show $Q^{(d)}(k)$ for different fitness distributions: (a) shows different configurations with fitness distributed exponentially on the unit-interval, with average fitness $\langle x \rangle = 0.05$. (b) shows different configurations with Pareto distributed fitness on the interval $[10^{-3}, 1]$, whereby the zero is excluded to avoid the singularity. Random selection, i.e $f(x, y) = 1$ is marked with hollow symbols, mutual attractiveness, i.e. $f(x, y) = xy$ with solid markers. The different plot markers, distinguish different average degrees: $\langle k \rangle = 1$ squares (yellow), $\langle k \rangle = 2$ circles (blue), $\langle k \rangle = 5$ triangles (brown), $\langle k \rangle = 10$ downward triangles (purple). The lines indicate analytical results from Eq. (4.19), whereby solid lines represent the case of $f(x, y) = xy$ and dashed lines $f(x, y) = 1$. Numerical results are obtained in a network with $N = 10^4$ nodes.

The conditional expectation of the node-degree is

$$\mathbb{E}[k|x] = \sum_{k \geq 0} k p^{(d)}(k|x) = 1 + \theta_q(x), \tag{4.20}$$

and the average degree

$$\langle k \rangle = \int_0^1 \mathbb{E}[k|x]\rho(x)dx = 2/q. \tag{4.21}$$

As in the previous section, this can be verified by considering that the average degree must equal twice the number of edges divided by the number of nodes in the network. This fraction is for $t \to \infty$ given by $2t/qt = 2/q$. Fig. 4.3 shows the comparison of numerical simulations to the analytical prediction of Eq. (4.19). The agreement between them is excellent. Compared to Fig. 4.1, the probability $Q^{(d)}(k)$ is broader than $Q^{(s)}(k)$. Thus vulnerable nodes with high degrees are more likely to exist in the dynamic model than in the static model. Another interesting graphical indicator is the degree-conditional probability that a randomly chosen node is vulnerable. This quantity corrects for implications of the degree distribution. Fig. 4.4a displays this probability, given by $Q^{(d)}(k)/P^{(d)}(k)$. Compared to Fig. 4.2a, the decay of $Q^{(d)}(k)/P^{(d)}(k)$ occurs at smaller values of k for all different $\langle k \rangle$ configurations. Also, the decay is slower than in the static case and largely influenced by different values of $\langle k \rangle$. Qualitatively, Figs. 4.4a and 4.2a are similar, they both display the phenomenon of adverse diversification, that is discussed above. However, also in the
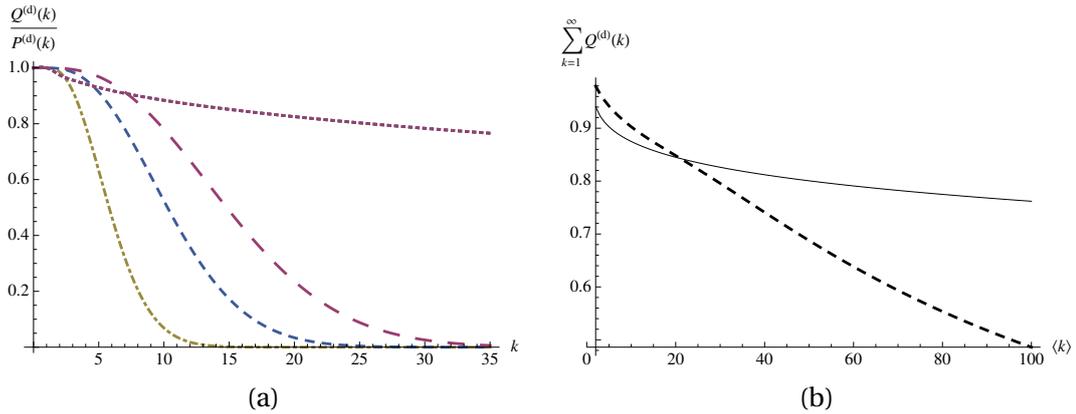
Figure 4.4: Plot (a) shows degree-conditional probabilities of vulnerability for different average degrees $\langle k \rangle = 2$ dot-dashed line (mustard), $\langle k \rangle = 5$ fine dashed (blue), $\langle k \rangle = 10$ coarse dashed (purple). The dotted line is an overlap of all $\langle k \rangle$ configurations with random attachment, i.e. $f(x,y) = 1$. All other lines correspond to $f(x,y) = xy$. The underlying fitness distribution is a power-law with $\langle x \rangle = 0.05$. The overlapping lines indicate that for random attachment, the conditional probability to find a vulnerable node is independent of the average degree $\langle k \rangle$. Plot (b) illustrates the relationship between the fraction of vulnerable nodes in the network and the average degree. The underlying configuration is again a power-law on $[10^{-3}, 1]$ with $\langle x \rangle = 0.05$. Solid (dashed) line corresponds to $f(x,y) = xy$ ($f(x,y) = 1$).

dynamic case is the overall fraction of vulnerable nodes decaying in increasing $\langle k \rangle$, see Fig. 4.4b.

3   Percolation

After the main topological property of interest, the degree distribution of vulnerable nodes, has been derived in the previous section, this section reviews the methodology that is used to calculate the main properties of the vulnerable component. The approach that is taken here has been used elsewhere [87, 91, 92, 124] and is reviewed for completeness.

First, define the generating function of the probability that a given node has degree k and is vulnerable $Q^{(i)}(k)$, as

$$G_0^{(i)}(s) = \sum_{k \geq 0} Q^{(i)}(k) s^k, \quad i \in \{\text{s,d}\}. \tag{4.22}$$

The superscript notation that indicates the specific model is suppressed in the following. Furthermore, the excess degree distribution [92] describes the probability that a

randomly chosen neighbour of a randomly picked node is vulnerable and has $k+1$ neighbours in total (k many neighbours, without the randomly picked node). To be more precise, it is the probability to find a vulnerable node with degree $k+1$, following a randomly chosen edge. This probability is proportional to the number of edges that are adjacent to vulnerable nodes with degree $k+1$, correctly normalised with the average degree in the network:

$$R(k) = \frac{(k+1)Q(k+1)}{\sum_{k \geq 0} k P(k)}. \tag{4.23}$$

The normalization is correct, since the node is a random choice taken from the set of all nodes, and not just from the set of vulnerable ones. Also, the generating function for $R(k)$ is defined as

$$G_1(s) = \sum_{j \geq 0} \frac{(j+1)Q(j+1)s^j}{\langle k \rangle} = \frac{G_0'(s)}{\langle k \rangle}. \tag{4.24}$$

Denote the number of nodes that can be reached following only connected nodes along their edges with t and the distribution of t with $\phi(t)$. The corresponding generating function is defined as

$$H_1(x) = \sum_{t \geq 1} x^t \phi(t). \tag{4.25}$$

To find an expression for $H_1(x)$, define the probability that a node with degree k belongs to a vulnerable component of size c as $\zeta(c|k)$. $\zeta(c|k)$ can be derived constructively by noting that the sum of nodes that can be reached following each of the nodes edges must sum up to $c-1$:

$$\zeta(c|k) = \sum_{t_1 \geq 1} \cdots \sum_{t_k \geq 1} \delta \left(c-1, \sum_{m=1}^{k} t_m \right) \prod_{m=1}^{k} \phi(t_m) \tag{4.26}$$

Where $\delta(x, y)$ is the Kronecker Delta. Now denote the probability that a randomly chosen node belongs to a vulnerable component of size c with $\pi(c)$. This is simply

$$\pi(c) = \sum_{k \geq 1} Q(k)\zeta(c|k). \tag{4.27}$$

The generating function for $\pi(c)$ can be computed in the same way as for example out-

lined in [92]:

$$\sum_{c\geq 1}\pi(c)x^c = \sum_{c\geq 1}x^c\sum_{k\geq 1}Q(k)\zeta(c|k) \tag{4.28}$$

$$= \sum_{k\geq 0}Q(k)\sum_{c\geq 1}x^c\sum_{t_1\geq 1}\cdots\sum_{t_k\geq 1}\delta\left(c-1,\sum_{m=1}^{k}t_m\right)\prod_{m=1}^{k}\phi(t_m) \tag{4.29}$$

$$= x\sum_{k\geq 0}Q(k)\sum_{t_1\geq 1}\cdots\sum_{t_k\geq 1}x^{\sum_{m=1}^{k}t_m}\prod_{m=1}^{k}\phi(t_m) \tag{4.30}$$

$$= x\sum_{k\geq 1}Q(k)\left(\sum_{t\geq 1}\phi(t)x^t\right)^k \tag{4.31}$$

$$= x\sum_{k\geq 1}Q(k)(H_1(x))^k \tag{4.32}$$

The result of the calculation above is the generating function of a vulnerable component of size one or greater. However, the possibility that the initially chosen node is not vulnerable, hence the random choice selected an empty vulnerable component of size zero must also be taken into account. This probability is given by $1-\sum_{k\geq 0}Q(k)$, hence the generating function for the probability that a randomly chosen node is part of a vulnerable component is given by

$$H_0(x) = 1 - G_0(1) + xG_0(H_1(x)). \tag{4.33}$$

The generating function for $H_1(x)$ can be established in a similar way and is given by

$$H_1(x) = 1 - G_1(1) + xG_1(H_1(x)). \tag{4.34}$$

The form of these two equations is standard. Equivalent results can also be found in [25, 124] for example.

The average size of the vulnerable component can be obtained from $H_0'(1)$ as

$$\langle n\rangle = H_0'(1) = \frac{G_0(1)+(G_0'(1))^2}{\langle k\rangle - G_0''(1)}. \tag{4.35}$$

The phase transition between a finitely sized vulnerable component and a vulnerable component that spans over the entire network can be calculated from Eq. (4.35). The

infinitely sized cluster emerges when $\langle n \rangle$ diverges, thus when $\langle k \rangle = G_0''(1)$. This condition can be written for the static model as

$$\sum_{k \geq 0} \frac{k(k-1)}{\Gamma(k+1)} \int_0^{1/k} e^{-M/N \lambda(x)} \left(\frac{M}{N} \lambda(x) \right)^k \rho(x) dx = \frac{2M}{N}. \tag{4.36}$$

For the dynamic model it is given by

$$\sum_{k \geq 0} k(k-1) \int_0^{1/k} \left(\frac{\theta_q(x)}{1 + \theta_q(x)} \right)^{k-1} \frac{\rho(x) dx}{1 + \theta_q(x)} = \frac{2}{q}. \tag{4.37}$$

Eqs. (4.36) and (4.37) define the percolation threshold, in other words for any two of $\rho(x)$, $\langle k \rangle = 2M/N$, $\lambda(x)$, respectively $\rho(x)$, $\langle k \rangle = 2q$ and $\theta_q(x)$ held fixed, these equations define the point of phase transition between an infinite number of small components and one giant vulnerable component in terms of the third variable.

Apart from the percolation threshold, other quantities of interest can also be derived from the theory that is laid out above [25, 87, 124]. Eq. (4.33) describes the generating function for the sizes of vulnerable clusters in the network. This generating function is expressed in terms of two other generating functions: $G_0(1)$, which is already calculated above, and $H_1(x)$ which is given implicitly in Eq. (4.34). Remember, $H_0(x)$ is the generating function for the distribution of outbreaks outside the percolating cluster [25]. The fraction of nodes in the largest vulnerable component S can therefore be computed with $H_0(1)$:

$$S = 1 - H_0(1). \tag{4.38}$$

S can be computed using Eq. (4.33) and solving Eq. (4.34) numerically [25]:

$$S = G_0(1) - G_0(\xi), \tag{4.39}$$

whereby ξ solves

$$\xi = 1 - G_1(1) + G_1(\xi). \tag{4.40}$$

One trivial solution of Eq. (4.40) is $\xi = 1$. The function $\varphi(\xi) = G_1(\xi) - \xi + 1 - G(1)$ can maximally have one more root. This follows from the following consideration. The derivative

of $\varphi(\xi)$ is $G_1'(\xi) - 1$ and the derivative of the derivative is $G_1''(\xi)$ which is strictly positive by definition of $G_1(\xi)$. Thus the derivative of $\varphi(\xi)$ must be a strictly increasing function. If a second root exists the derivative must be zero somewhere $\in (0, 1)$, moreover positive at $\xi = 1$. This condition translates into

$$G_1'(1) - 1 > 0. \tag{4.41}$$

This concept has been illustrated graphically in Fig. 2.9 in chapter 2. Substituting the definition of $G_1(x)$ into Eq. (4.41) leads to

$$G_0''(1) > \langle k \rangle. \tag{4.42}$$

As expected, this condition is equivalent to the percolation condition given in Eqs. (4.36) and (4.37). Therefore below percolation, the only solution to Eq. (4.40) is $\xi = 1$, which translates into $S = 0$, which makes sense since S is the size of the giant component which does not exist below percolation.

4   Results

In this section different attachment kernels and two different density distributions are considered. These are mutual and random attachment i.e. $f(x, y) = xy$ and $f(x, y) = 1$, coupled with power-law fitness $\rho(x) \propto x^{-a}$, and $\rho(x) \propto e^{-ax}$. The two specific fitness distributions have been chosen since they allow an interesting comparison between various network motifs. Power-law probability distributions can be found in many areas of scientific interest [3, 31]. But also exponential distributions are justifiable in certain scenarios and can be found in real data [43]. Exponential distributions are not at last also considered to illustrate the effect of heavier tails on percolation in this class of networks. While the theory in the main part of this chapter is derived for fitness distributed over the unit interval, the domain of the power-law is constrained to $[10^{-3}, 1]$ for the numerical evaluation in order to avoid the singularity at $x = 0$. The form of the resulting degree distribution of these networks follow two distinct patterns, power-law and exponential.
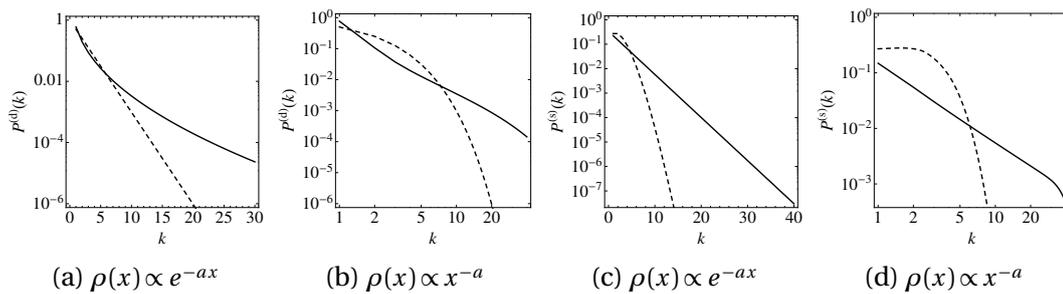
(a) $\rho(x) \propto e^{-ax}$ (b) $\rho(x) \propto x^{-a}$ (c) $\rho(x) \propto e^{-ax}$ (d) $\rho(x) \propto x^{-a}$

Figure 4.5: Degree distributions for the different configurations that are under consideration. Dashed lines for random attachmentment $f(x,y) = 1$ and solid lines for mutual attractiveness $f(x,y) = xy$ with $\langle x \rangle = 0.05$ and $\langle k \rangle = 2$ in all cases. (a) Dynamic model, exponential fitness distribution. (b) Dynamic model, power-law fitness. (c) Static model, exponential fitness, (d) Static model, power-law fitness.

This is illustrated in Fig. 4.5. It is clear from Fig. 4.5 that most configurations lead to an exponential or exponentially decaying degree distribution. However, as is elucidated below, the resulting spreading behaviour is entirely different from case to case. Notice as well the different behaviour between the dynamic and the static fitness model. While the dynamic model produces an almost perfect power-law degree distribution over the whole range of k for $f(x,y) = xy$ and power-law fitness, the static model produces a power-law with cutoff at large k. The finding that non of the distributions leads to a power-law over the entire domain can be explained by the bounded fitness interval. In order to produce a power-law within this regime, an attractor $\lambda(x)$, respectively $\theta_q(x)$ with a singularity at $x = 1$ is necessary. This has been elaborated in chapter 3. The fact that many configurations lead to similar degree distributions, but display an entirely different failure spreading behaviour highlights the importance of the distinction between static and dynamic fitness models.

Fig. 4.6 illustrates analytical results on the vulnerable fraction S depending on the average fitness $\langle x \rangle$ and average degree $\langle k \rangle$. Results for the dynamic and the static model are displayed separately. The differences between these two models are not only quantitative, but also of qualitative nature. Notice that the occupied fraction S as defined in Eq. (4.39) is bound from above by $G_0(1) = \sum_{k \geq 0} Q(k)$. Since $Q(k)$ is normalised to the total number of nodes, S can possibly be bound from above at less than 1. This is one of reasons why the graphs in Fig. 4.6 appear to be cut off. Another reasons is that for sufficiently low $\langle x \rangle$ the connectivity of the possibly percolating component collapses and the component vanishes. For the case mutual attractiveness (solid characters), the network which
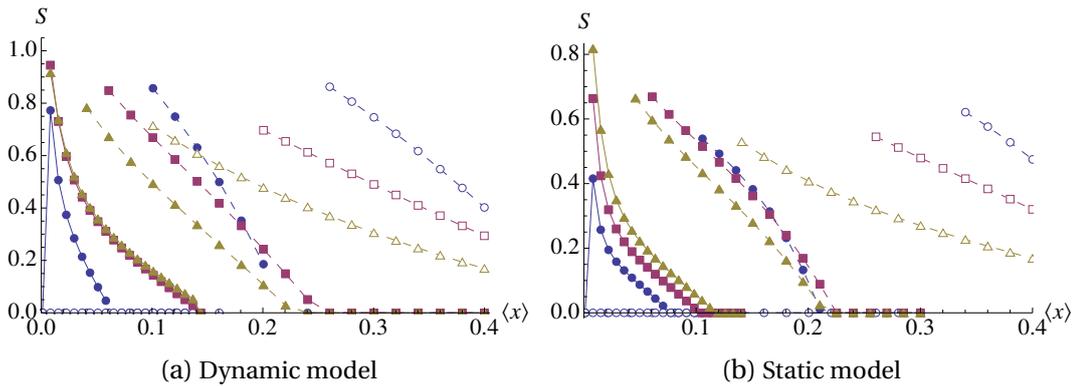
(a) Dynamic model   (b) Static model

Figure 4.6: Occupied fraction of the percolating vulnerable component in the dynamic (a) and static (b) network. The different plot character indicate different average degrees: ○ (blue) – $\langle k \rangle = 2$, □ (magenta) – $\langle k \rangle = 5$, and △ (mustard) – $\langle k \rangle = 10$ The different attachment kernels are indicated with solid plot characters for $f(x,y) = xy$ and hollow characters for $f(x,y) = 1$. Finally, the two different fitness distributions are distinguished by line type. Solid line for $\rho(x) \propto x^{-a}$ and dashed line for $\rho(x) \propto e^{-ax}$.

is induced by power-law fitness (solid line) is significantly more robust than its exponentially induced counterpart (dashed line). The onset of the epidemic occurs at much larger $\langle x \rangle$ for the latter case. Curiously, the effect of differing average degree depends on the exact configuration. Generally, the higher the average degree, the lower is the size of the vulnerable component. This is intuitive, since the vulnerability condition is $x < 1/k$ and for larger k this condition is harder to fulfil. However, for the case of mutual selection and power-law fitness (solid characters, solid line) the vulnerable component has the smallest size for $\langle k \rangle = 2$ (blue circles). This illustrates that the structure of the underlying network plays an important role and has a non-trivial effect on the connectivity of the vulnerable component. Another configuration where this becomes clear is mutual attractiveness (solid characters) and exponential fitness (dashed line). The graphs of S for $\langle k \rangle = 2$ (blue circles) and $\langle k \rangle = 5$ (magenta squares) intersect. This is another aspect that highlights the non-trivial connectivity of the percolating component. The case of random attachment (hollow characters) and exponential fitness (dashed line) is special. Global connectivity of the percolating component sets on at large values of $\langle x \rangle$ compared to the other graphs, and the decay is very slow. Thus, this network is hard to stabilise once the percolating component can reach global connectivity. Apart from the individual realisations of $S(\langle x \rangle)$, Fig. 4.6 illustrates that there are two competing effects at work. On one side, higher average fitness stabilises the system, since nodes with higher fitness are more resilient. This effect induces the decay towards higher $\langle x \rangle$ regions. On

the other side, when the average fitness falls below a certain value, the connectivity of low fitness nodes that are possibly prone to fail is not sufficient to build a global failing component which leads to the abrupt cutoffs on the left-hand side of the graphs.

5   Conclusions

The coupling of propensity to form edges in the network with robustness against failures using one variable called fitness has been discussed in this chapter. The assumption of this coupling is realistic in many scenarios because the robustness – be it financial or health – is directly related to the attractiveness of an individual in a contact process. The resulting theory can help to devise immunisation strategies not through direct rewiring or protecting single nodes but rather by incentivising individuals to form a more stable structure. It is shown in this chapter that the introduction of a tax that changes the distribution of fitness in the system or the inception of a stimulus to form new edges within the network can stabilise the network without violating individual preferences that are described by $f(x, y)$. Another aspect that has been illustrated in this chapter is the importance of the distinction between dynamic and static fitness models. The percolation behaviour of these two network classes is has been found to be different. Also, the broadness of the fitness distribution has an effect on the stability of the network. A higher variety of fitness among individuals stabilises the network. Moreover, the attachment kernel has a profound influence on the way an epidemic spreads. Mutual attractiveness, i.e. $f(x, y) = xy$ induces a network that is significantly more resilient than a network composed by pure random attachment.

In this chapter, the networks under investigation were flat so that shocks could only propagate on one layer. For future research, it would be interesting to investigate how a multiplex architecture with similar rules behaves and whether the findings are comparable to ones that were found here.

Chapter 5

The World Trade Network

Previous studies have suggested that the world-trade network belongs to the class of static hidden variable models. In this chapter, the microscopic structure of the world trade network, that is the hidden variable correlation matrix of the network, is investigated. The hidden variable is defined as a rank ordering of gross domestic products. This choice significantly reduces the noise in the statistical analysis found in previous studies. The hidden variable correlation matrix, that expresses the probability that a trade relationship between two countries of given fitness exists, suggests an attachment kernel that at least partially favours trading pairs or dissimilar fitness rather than the purely multiplicative one found previously. Additionally, an in-depth look at the data source is provided and it is revealed that first-order results, such as the degree distribution, exhibit significant qualitative differences depending on the data provider. Furthermore, light is shed on the intertemporal activity of international trade and it is pointed out that fluctuations occur mostly between countries with strong dissimilarities of fitness and connectivity.

1   Introduction

The world trade network (WTN) has received a considerable amount of attention in the last decade. The detailed understanding of the network structure of international trade facilitates a better comprehension of crises that spread across a country's frontiers. It is generally agreed that a static hidden variable model [24, 111] best describes the topology of the WTN. The static model is most suitable because there are only a finite number of distinct countries in the world. Depending on the scope of the analysis, previous contributions understand the network as growing since the number of nodes increased from 86 in the year 1950 to 190 in the year 2000 [56]. However, since there are –depending on the definition– approximately 195 countries in the world to date, the growth of the network cannot be sustained. It has been shown in [57, 58], that the static fitness model is an adequate model to describe the structure of the world trade network for several years, not just on the level of node degree distributions, but up to higher order statistics, such as the average nearest neighbour degree. In this chapter, a view into the microscopic structure of the WTN is provided. The main quantity of interest is the edge density, that is the probability matrix for any two nodes of fitness x and y to share an edge. The fitness model is found to be an excellent fit for the WTN. However, the previously found multiplicative attachment kernel varies significantly from the kernel that is found here by investigating the edge density.

A review of the literature on the network of international trade reveals that many results are not without controversy. Even first order results, such as the form of the degree distribution, differ between various publications. One reason for these discrepancies is the choice of data source. Every trade relationship between two countries should be reported twice, once as an export and once as an import. This is however not always the case [53]. As a result, different data providers report different trading data. Another issue is the way these disparities are corrected. Some [10, 115] simply use only import data, which is deemed more reliable than export data, others [18, 51, 54] interpolate between reports of exports and imports.

The data set that is used for this chapter is provided by the National Bureau of Economic Research (NBER) [53] and forms an excellent starting point for this investigation. The special feature of this data set is that one part of it has been adjusted for errors, while another part has not. The adjustments have been made by specialised macroeconomists and can therefore be deemed reliable.

The world trade network is treated as directed and binary in this chapter. That means that directionality of trade is taken into account, but the traded amounts are ignored. Others [49, 51, 58, 109] have argued that directionality could be ignored because of the high reciprocity, i.e. export links are usually reciprocated with import links. However, since the world trade network is directed by nature, the analysis in this chapter is laid out in terms of a directed network. A weighted representation of the WTN has been investigated for example in [18, 50, 51]. Curiously, the additional information of edge weight is not always of greater explanatory power than binary network analysis [115, 116]. In [115, 116] a network randomisation technique was used to show that by knowing the degree sequence of the binary WTN, higher order statistics like the average neighbour degree and the clustering coefficient can be obtained. However, using a weighted network approach and fixing the sequence of interaction strengths, these higher order statistics cannot be found anymore. This suggests that the binary network analysis that is used in economics is a powerful tool, because all the necessary information is contained in the first order statistics.

In principle, the WTN is a spatially embedded network. However, it has been illustrated, that distances between countries do not add a significant amount of information to a binary analysis of international trade [101]. Similarly, also in the weighted regime, the importance of geographic distance for understanding international trade is declining over time [30]. The results of [30] are in contrast to the standard notion in economics, that the intensity of trade between pairs of countries is strongly related to their distance, see for instance [79]. Another aspect that will not feed into the analysis in this chapter is the multi-layered architecture of international trade. Every reported trade flow is an aggregation over different product categories. Refs. [10, 11, 115, 116] investigate these different layers separately.

The investigation of the topology of the world trade is not conducted as an end in itself. Its aim is to understand how trade can affect economic welfare. Many aspects are yet to be understood. The theory of complex systems is just one of the building blocks towards a good understanding of those effects [107, 108]. Network-theoretic measures have been shown to explain parts of nation's income. In [73] it is shown that an improvement of the degree centrality ranking by ten units, increases the average GDP per capita by 0.27%. Others [54, 56, 57, 59] investigate correlation structures of income, connectivity and interaction strength in weighted networks and illustrate that an involvement in international trade has a direct impact on income and vice versa. Additionally, it has been shown that network properties have good explanatory power to detect vulnerable economies in the WTN [55, 74].

This chapter is organised as follows: in section 2, the static fitness model is reviewed and the central quantities for this study derived. In section 3, the data set is introduced and reasons for its choice are discussed. In section 4, the definition of fitness is elucidated. In section 5, the inter-temporal structure of the WTN is clarified and in section 6, the static structure is investigated. Section 7 closes the chapter with concluding remarks.

2 The Static Fitness Model

The empirical analysis in this chapter is based on the static fitness model, as it was introduced in [21, 24, 111]. The investigation that is presented later in the text relies on results that are reviewed in the following.

The static fitness model is a network model with N nodes and M directed edges. The formalism is the same as in the previous chapters: each node inside the network is endowed with a fitness value x, that is drawn from a probability density function $\rho(x)$. The probability that a node with fitness x originates a link toward a node with fitness y is proportional to the attachment kernel $f(x, y)$. Although the network is static, the edge deployment process can be understood as dynamic. The probability that an edge is added to a pair of nodes $i \rightarrow j$ is $f(x_i, x_j)/\sum_{k,l} f(x_k, x_l)$. It should be noted here that this for-

malism ignores the possible problem of edge duplications. Edge duplications are not an issue in sparse networks, i.e. $M \ll N^2$, however they will lead to bias in dense networks, such as the WTN. However, correcting for the impossibility of duplications leads to non-linearities in the formalism, so that closed form solutions can no longer be found. The numerical results that are presented later in this chapter illustrate that this bias is sufficiently small and the introduction of non-linear terms in the formalism is not necessary for the scope of this analysis.

To begin with, the probability that two nodes with fitness x_i and x_j are connected in a network with M edges and N nodes is derived. Denote this probability with $C_{M,N}(x_i, x_j)$ and further define $\overline{C}_{M,N}(x_i, x_j) = 1 - C_{M,N}(x_i, x_j)$. $C_{M,N}$ will be referred to as the microscopic structure of this network because it contains all information on the network and thus all topological properties of the network can be derived from it. Assuming a sequential update of the network, $C_{M,N}$ can be derived from first principles using

$$\overline{C}_{M+1,N}(x_i, x_j) = \overline{C}_{M,N}(x_i, x_j) \left(1 - \frac{f(x_i, x_j)}{\sum_{k,l} f(x_k, x_l)} \right). \tag{5.1}$$

That is, the probability that there exists no edge $i \to j$ in a network with $M+1$ edges is equal to the probability that there was no such edge in a network with M edges and no edge has been added to the pair $i \to j$. Assuming now sufficiently large N, such that fitness can be approximated as continuous variable leads to

$$\overline{C}_{M+1,N}(x, y) = \overline{C}_{M,N}(x, y) \left(1 - \frac{f(x, y)}{N^2 \int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta} \right). \tag{5.2}$$

Using the boundary condition $\overline{C}_{0,N}(x, y) = 1$ and a continuous time approximation, the solution of Eq. (5.2) is given by

$$\overline{C}_{M,N}(x, y) = \exp \left\{ -\frac{M}{N^2} \frac{f(x, y)}{\int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta} \right\}. \tag{5.3}$$

This expression can be approximated, assuming a sufficiently sparse network, such that

the probability for the existence of an edge is given by

$$C_{M,N}(x,y) \cong \frac{M}{N^2} \frac{f(x,y)}{\int_0^1 \int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta}. \tag{5.4}$$

Another quantity of interest in this chapter is the expected in-degree of a node with a given fitness value. This quantity can also be derived from first principles. Define the probability that a node with fitness x increases its in-degree in an edge-addition step with

$$\lambda(x,N) = \frac{1}{N} \frac{\int_0^1 f(y,x)\rho(y)dy}{\int_0^1 \int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta}. \tag{5.5}$$

Since the size of the network N is a factor of $\lambda(x,N)$, also define $\lambda(x) = N\lambda(x,N)$. Notice that this definition of $\lambda(x,N)$ differs from the definition in the previous chapter just by the factor 2, where the network has been assumed to be undirected. Furthermore, $p_{M,N}(k|x)$ is the probability that a node with fitness x has in-degree k in a network with M edges and N nodes. $p_{M,N}(k|x)$ obeys

$$p_{M+1,N}(k|x) = p_{M,N}(k|x)(1 - \lambda(x,N)) + p_{M,N}(k-1|x)\lambda(x,N). \tag{5.6}$$

This equation can be solved using a generating function approach. The resulting conditional in-degree distribution is

$$p_{M,N}(k|x) = \frac{e^{-M/N\lambda(x)}\left(\frac{M}{N}\lambda(x)\right)^k}{\Gamma(k+1)}. \tag{5.7}$$

Where $\Gamma(x)$ is the standard Gamma function. This result is the same as the one found in Eq. (4.7), apart from the different definition of $\lambda(x,N)$ to account for directionality. Notice that the out-degree distribution can be described in a very similar way. The only part that changes is the attractor $\lambda(x,N)$. The probability that a node increases its out-degree by one in an edge addition step is

$$\mu(x,N) = \frac{1}{N} \frac{\int_0^1 f(x,y)\rho(y)dy}{\int_0^1 \int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta}. \tag{5.8}$$

Thus if $f(x,y)$ is a symmetric function, then the conditional out-degree- and in-degree

distributions are identical. The fitness conditional expectation of the in-degree of a randomly chosen node is

$$\bar{k}(x) = \mathbb{E}[k|x] = \sum_{k \geq 0} k\, p_{M,N}(k|x) = \frac{M}{N}\lambda(x). \tag{5.9}$$

Thus, the form of the conditional expectation is solely determined by the form of λ. The quantities $C_{M,N}(x,y)$ and $\bar{k}(x)$ are closely related, more specifically:

$$\int_0^1 C_{M,N}(x,y)\rho(x)dx = \frac{M}{N^2} \frac{\int_0^1 f(x,y)\rho(x)dx}{\int_0^1 \int_0^1 f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta} \tag{5.10}$$

$$= \frac{M}{N^2}\lambda(y) \tag{5.11}$$

$$= \frac{1}{N}\bar{k}(y). \tag{5.12}$$

This relationship will be used in a later section to assess the validity of the fitness model for the WTN.

3 The Choice of a Data Source

The choice of the data set is crucial for the validity of the derived results. Data for the WTN can be found in various publications [53, 60, 94]. The concern of data validity is raised in various places [10, 18, 51, 53, 115]. Data on world trade should always be reported in two statements. Every flow of goods or services occurs in one country as an export and in another country as import. However, it occurs that a significant amount of trade is only reported on one side. This leads to inconsistencies in the data analysis. Some suggest methods to approximate the true value of money flow by averaging over the existing data [18, 51], while others [10, 115] suggest to replace all export data with import data, because import data is believed to be more reliable [53].

A technique to correct for the inconsistencies is not suggested here, merely the severity of the problem is pointed out. The network data for this study is obtained from the National Bureau of Economic Research (NBER). A detailed description is given in [53]. The
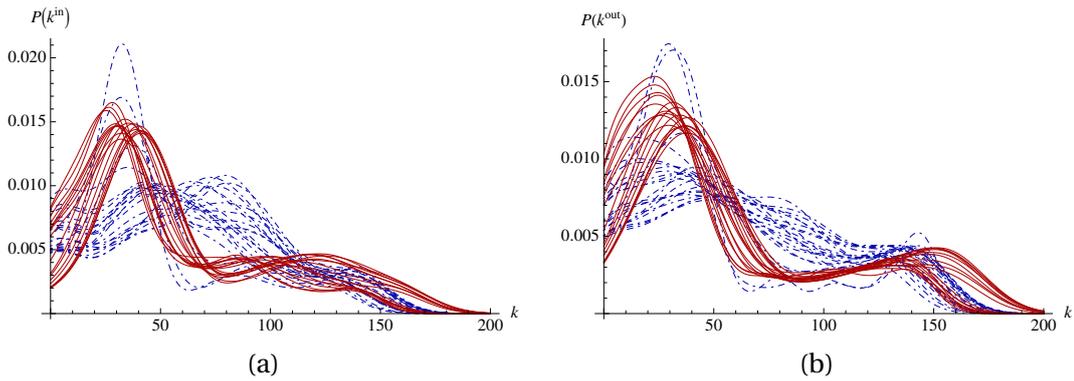
Figure 5.1: Density estimators for the (a) in-degree and (b) out-degree distribution in the world trade network for the years 1962-2000. The data before the year 1984 is indicated with dot-dashed lines (blue), the distributions for the years 1984 and after is plotted solid red. The two outliers in the sample of graphs for the years before 1984 in (b) represent the years 1978 and 1983.

primary data source for [53] is two-fold. Data for the years 1962-1983 is taken from UN data, collected at various times and has not been revised during later stages. Data from 1984-2000 is also taken from UN Comtrade data, but has undergone significant revisions and corrections that are explained in detail in [53]. The corrections can be considered reliable, as experts in the field have searched for data and corrected it manually. The corrections were conducted at component level, i.e. each commodity group was considered in isolation.

The effect of this data adjustment is illustrated in Fig. 5.1. The figure shows density estimators for the in- and out-degree distribution of the WTN over the years 1962-2000. The two colours indicate the separation into data that was curated and data that has undergone only little adjustment, if any. It is evident that the year 1984 marks a change of regime, which is most likely not inherent in the network formation process, but in the way the data is collected and revised, which has changed in that year. Comprehension of stochastic processes that take place on networks is often formulated in terms of the degree distribution [13]. It is therefore of high importance for further analysis of the network to have a correct representation of this quantity.

It is anticipated that a different data collection scheme will lead to slightly different empirical results, but Fig. 5.1 presents an entire regime shift of first-order results that is solely initiated by a change of the data source. All further results of this chapter rely on the relatively short, but well adjusted data that is discussed in [53] for the years 1984-

2000.

Next to the data on international trade, also data to resemble the hidden variables was collected. The gross-domestic-product (GDP) was obtained from UN-data [1]. The countries that are covered in this chapter are displayed in Tab. 5.1 at the end of the chapter.

4   Definition of Fitness

It is undisputed that the world trade network falls into the class of hidden variable models [56, 57, 58]. However, it is not clear how to define fitness. There is general agreement that fitness is defined over some constraint interval, usually $[0, 1]$. Two different definitions of fitness are investigated in the following.

One possibility is to impose a ranking on the node's GDP and normalise this ranking to the unit interval. Formally, denote the income of country i as w_i. Then the corresponding fitness of node i is defined as

$$x_i = \vartheta(w_i) = \frac{1}{N} \sum_{j=1}^{N} H(w_i - w_j)$$ (5.13)

Where $H(x)$ is the Heaviside step function. This definition is the discrete analogy to the ranking definition in chapter 3. Fig. 5.2 illustrates the relationship between fitness and expected node-degree. The almost linear form in the log-linear scaled plot implies an approximately exponential form of the fitness conditional expected node-degree $\mathbb{E}[k|x]$. This empirical finding will be compared later with analytical results from Section 2. Notice that the particular choice of fitness in Eq. (5.13) simplifies the expression of $\lambda(x)$ significantly. Suppose that the national income X is distributed with density $f_X(x)$. Consider fitness as a transformed variable, in the sense that $Y = \vartheta(X)$. The cumulative distri-
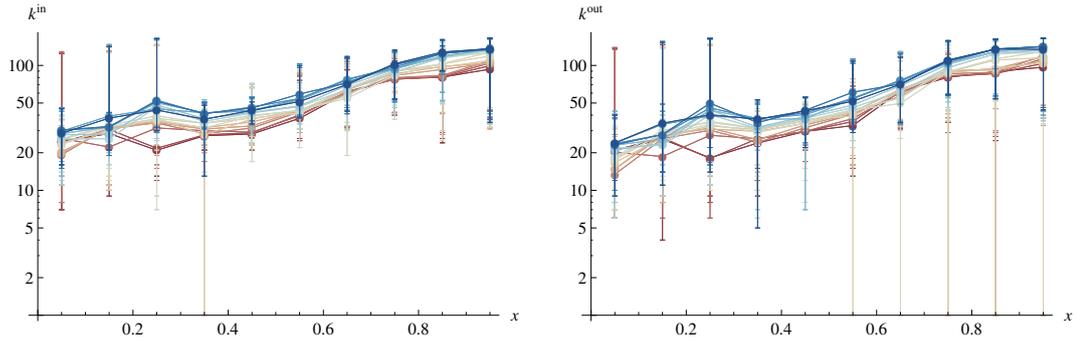
---

[1] http://www.data.un.org/

Figure 5.2: Relationship between the in-degree (a), respectively the out-degree (b) of a node with its fitness, when fitness is defined as in Eq. (5.13). The log-linear plots show an almost linear dependence which suggests that $\mathbb{E}[k|x]$ has an exponential form. The data is binned into intervals of length 0.1, the averages of these intervals are indicated with bulletpoints. The maximum and minimum of the intervals are shown as errorbars. The different colors represent different years starting from 1984 (red) until 2000 (blue).

bution of fitness can be written as

$$R_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[\vartheta(X) \leq y] \tag{5.14}$$

$$= \mathbb{P}[X \leq \vartheta^{-1}(y)] \tag{5.15}$$

$$= \int_0^{\vartheta^{-1}(y)} f_X(x)dx. \tag{5.16}$$

The definition of $\vartheta(x)$ is simply a discrete formulation of the cumulative distribution function of income X. Thus $\int_0^x f(\xi)d\xi = \vartheta(x)$ and therefore it follows directly from Eq. (5.16) that

$$R_Y(y) = y \quad \text{thus} \quad \rho(y) = 1. \tag{5.17}$$

The fact that the fitness density becomes a constant eliminates one degree of freedom in the definition of $\lambda(x)$.

However, the particular definition of fitness is arbitrary. Another form that was suggested in Garlaschelli and Loffredo [57] is

$$\vartheta'(w_i) = \frac{w_i}{\sum_{j=1}^N w_j}, \tag{5.18}$$

or alternatively in [58]

$$\vartheta''(w_i) = \frac{w_i}{N\sum_{j=1}^N w_j} = \frac{\vartheta'(w_i)}{N}. \tag{5.19}$$

The corresponding relationship between the degree and the fitness of a node, using the
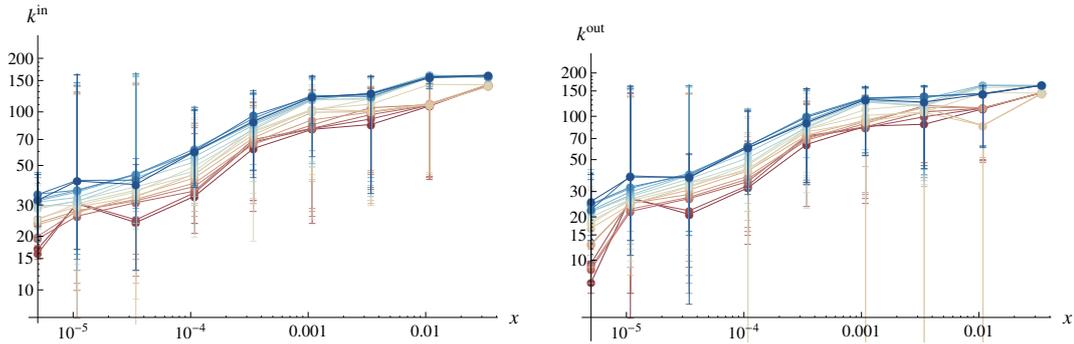
Figure 5.3: Relationship between the in-degree (a), respectively the out-degree (b) of a node with its fitness, when fitness is defined as in Eq. (5.18). The similar shape to the graphs in Fig. 5.2 shows that the expected degree is more stretched here, since the abscissa is logarithmically scaled, as opposed to the linear scaling in Fig. 5.2. The data is binned into logarithmic intervals, the averages of these intervals are indicated with bulletpoints. The maximum and minimum of the intervals are shown as errorbars. The different colors represent different years starting from 1984 (red) until 2000 (blue).

fitness definition in Eq. (5.18) is illustrated in Fig. 5.3. This sub-linear form in a log-log-scaled plot is the same as found in [57]. Notice that the empirical $\lambda(x)$ is slightly more noisy in the ϑ' regime than when fitness is defined by ϑ. However, the averages do not represent good approximations for the bins, the data in the bins is distributed over a wide range of k. The original derivation of various topological properties in [111] uses the fact that $\mathbb{E}[k|x]$ is invertible. That implies that the distribution of k inside the bins must be relatively narrow. Both fitness definitions seem to violate this assumption, however the results using ϑ appear narrower. For both definitions, there seems to be a more explicit relationship between fitness and the in-degree than between fitness and the out-degree.

5   Intertemporal Structure of the WTN

The structure of the world trade network is permanently changing, new trade relations are established and existing ones are terminated. Fig. 5.4 illustrates the macroscopic behaviour of the WTN over time. The number of nodes is almost static apart from the sudden increase of trading countries when the Soviet Union collapsed. Since this event is not system inherent, the network can be regarded as static with respect to the number of nodes. The number of edges is almost constantly increasing. In order to understand
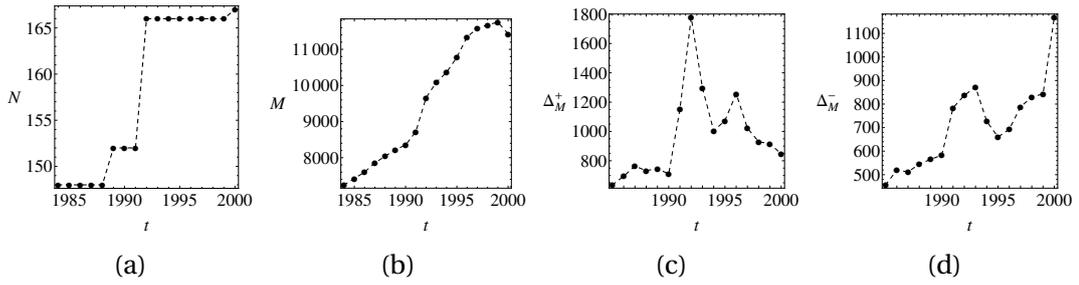
Figure 5.4: Macroscopic development of the WTN over time. (a) Temporal development of the number of nodes N. (b) Temporal development of the number of edges M. (c) Temporal development of the number of added edges as defined in Eq. (5.20). (d) Temporal development of the number of removed edges as defined in Eq. (5.20)

the change of the number of edges better, define

$$\Delta_M^{\pm}(t) = \sum_{i,j=1}^{N(t)} \delta[A_{ij}(t) - A_{ij}(t-1), \pm 1].$$ (5.20)

Where $\delta[k, k']$ is the Kronecker delta and $A_{ij}(t)$ is the adjacency matrix of the directed network at time t. Figs. 5.4 (c-d) reveal that the steady influx of new edges is composed by both edge deletion and dominating edge addition. Notice that the link density, defined as $M/(N(N-1))$ is steadily increasing over the reported time, here from 0.33... up to 0.41.... These figures remain the same in the case of the undirected network. This finding is in contrast to [18], where a roughly constant link density 0.52 over 53 years from 1948 to 2000 using an undirected network is reported. This difference illustrates very clearly the importance of a carefully chosen data source, as link density is a very coarse measure.

Apart from the macroscopic observation that edges constantly emerge and disappear, it is important to understand whether fitness is the main driver of these processes. Recently, activity driven networks have been investigated [100], see also the following chapter. Every node in this class of networks is endowed with an activity value that determines the probability that a particular node originates a new link inside the network. In order

to identify a coupling between activity in new trade links and fitness, define

$$e_{i,\pm}^{\text{out}}(t) = \sum_{j=1}^{N(t)} \delta[A_{ij}(t) - A_{ij}(t-1), \pm 1] \qquad \text{and} \tag{5.21}$$

$$e_{i,\pm}^{\text{in}}(t) = \sum_{j=1}^{N(t)} \delta[A_{ji}(t) - A_{ji}(t-1), \pm 1]. \tag{5.22}$$

$e_{i,\pm}^{\text{out}}(t)$ and $e_{i,\pm}^{\text{in}}(t)$ measure the number of edges that are added to node i either as outward or inward links. Activity levels can be measured as fractions of overall activity. Thus, define additionally

$$\zeta_{i,\pm}^{\text{out}}(t) = \frac{e_{i,\pm}^{\text{out}}(t)}{\sum_{j=1}^{N} e_{j,\pm}^{\text{out}}(t)} \quad \text{and} \tag{5.23}$$

$$\zeta_{i,\pm}^{\text{in}}(t) = \frac{e_{i,\pm}^{\text{in}}(t)}{\sum_{j=1}^{N} e_{j,\pm}^{\text{in}}(t)}. \tag{5.24}$$

Fig. 5.5 illustrates the dependency of fitness on $\zeta_{i,\pm}^{\text{in,out}}$ in a two-dimensional histogram. It is clear from Fig. 5.5 that fitness is not a good descriptor for activity levels, there is no clearly visible correlation between activity and fitness. Since no particular pattern emerges in Fig. 5.5, it is clear that a unilateral mechanism is not driving the intertemporal dynamics of the WTN as defined in Eqs. (5.23-5.24). Fig. 5.6 shows contour plots of the fitness and degrees of nodes adjacent to newly deployed and removed edges. Interestingly, two distinct pattern emerge. This shows that in contrast to the previously discussed activity histograms, a bilateral mechanism can explain the addition and removal well. Fig. 5.6 shows that new links emerge and disappear preferably between countries with dissimilar fitness and degree. The fitness contour plots (a), (c) for addition and removal of edges are qualitatively similar. Both have a high concentration around pairs of high fitness and further high densities for pairs of dissimilar fitness. Since this is the case not only for addition, but also for removals, it suggests that fitness does not only account for attractiveness, but also for activity in general. Curiously, the contour plots of node degrees adjacent to newly deployed and removed edges show an almost inverted pattern, compared to the fitness based contour plots. The fitness dependent plots show high activity mostly between high fitness nodes and nodes of differing fitness, while the
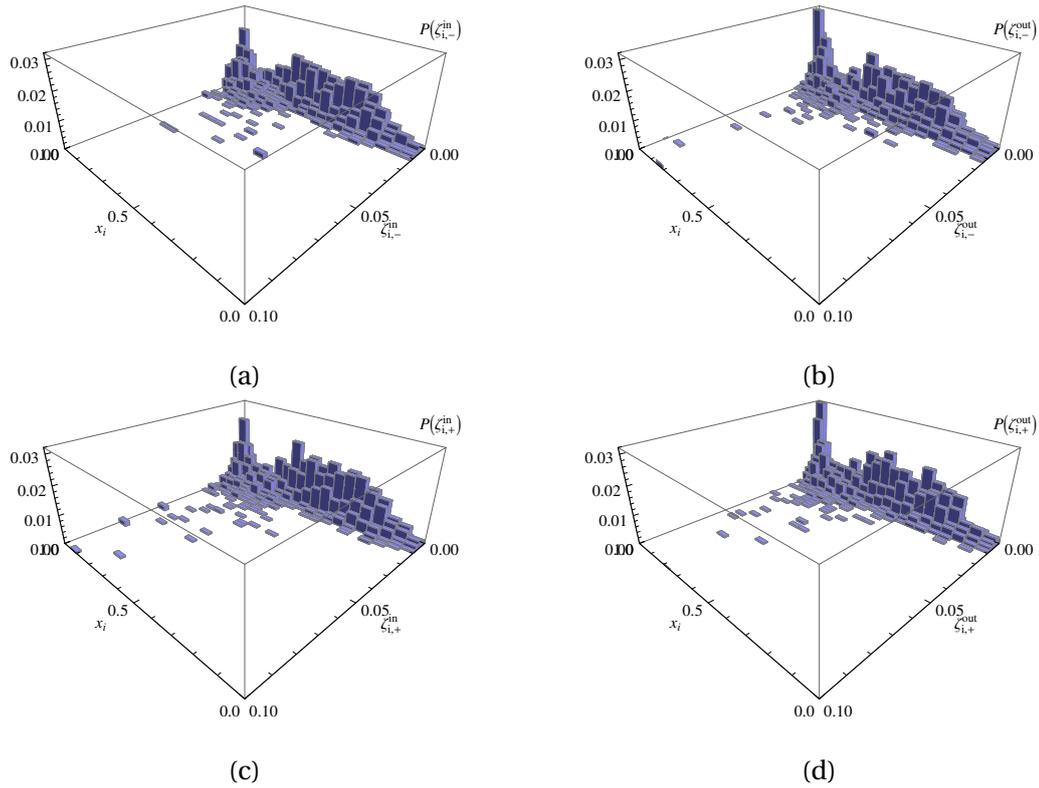
Figure 5.5: 2D-Histograms of activity levels as defined in Eqs. (5.23-5.24). (a) $\zeta_{i,-}^{\text{in}}$ vs. fitness. (b) $\zeta_{i,-}^{\text{out}}$ vs. fitness. (c) $\zeta_{i,+}^{\text{in}}$ vs. fitness. (d) $\zeta_{i,+}^{\text{out}}$ vs. fitness.

degree based plots show most of the activity with at least one low-degree node involved. This is surprising as it is pointed out earlier in this text that there is a strong positive relationship between fitness and degree. Another aspect that can be inferred from Fig. 5.6 is that new trade relationships emerge equally from low-connected to highly connected ones and vice versa. In other words, the attachment mechanism can assumed be to be symmetric.

6 Static Structure of the WTN

In order to get a more complete picture of the internal mechanism of the WTN, it is useful to look also at the mere existence of links, rather than at their emergence. The analysis here will be restricted to the years 1984 and 2000. Fig. 5.7 pictures the fitness and in-degrees of nodes that are adjacent to edges in the years 1984 and 2000. The results are similar to the findings in the previous section. Most of the trade flows between dis-

(a)

(b)

(c)

(d)

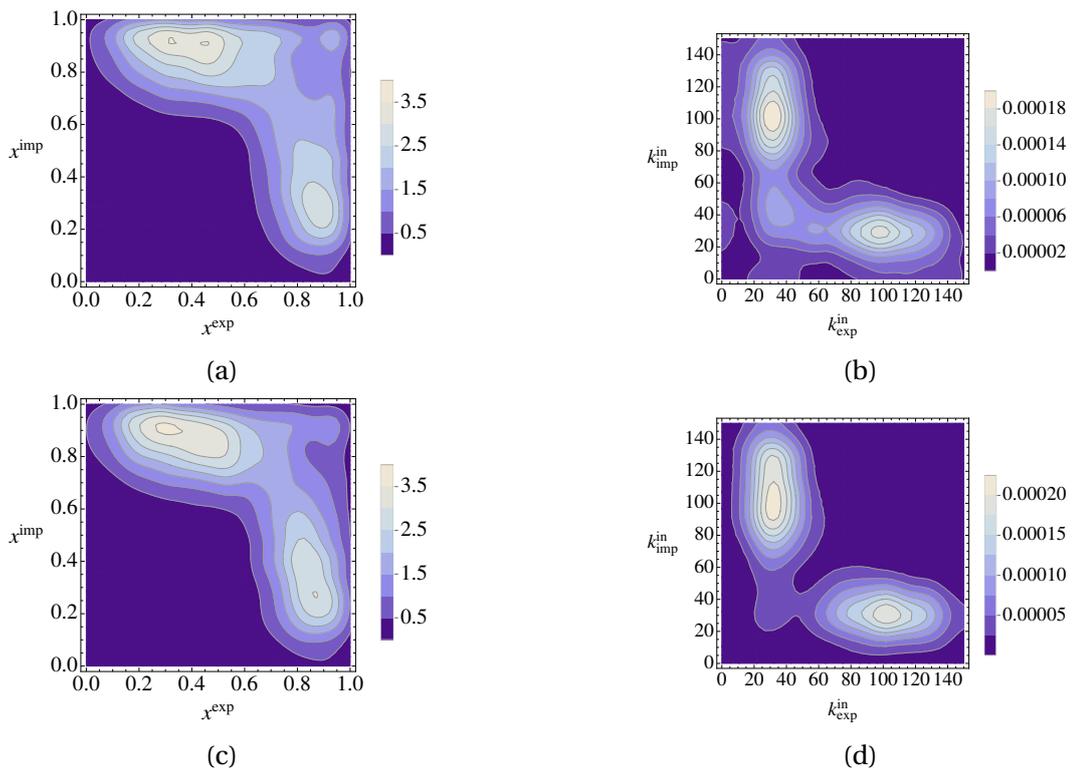Figure 5.6: Contour-plots of two dimensional histograms of in-degrees and fitness adjacent to removed and added edges. Dark areas indicate regions of low probabilities, white areas indicate high probabilities. (a) Fitness of nodes adjacent to newly deployed edges. (b) In-degrees of nodes adjacent to newly deployed edges. (c) Fitness of nodes adjacent to removed edges (d) In-degrees of nodes adjacent to removed edges.
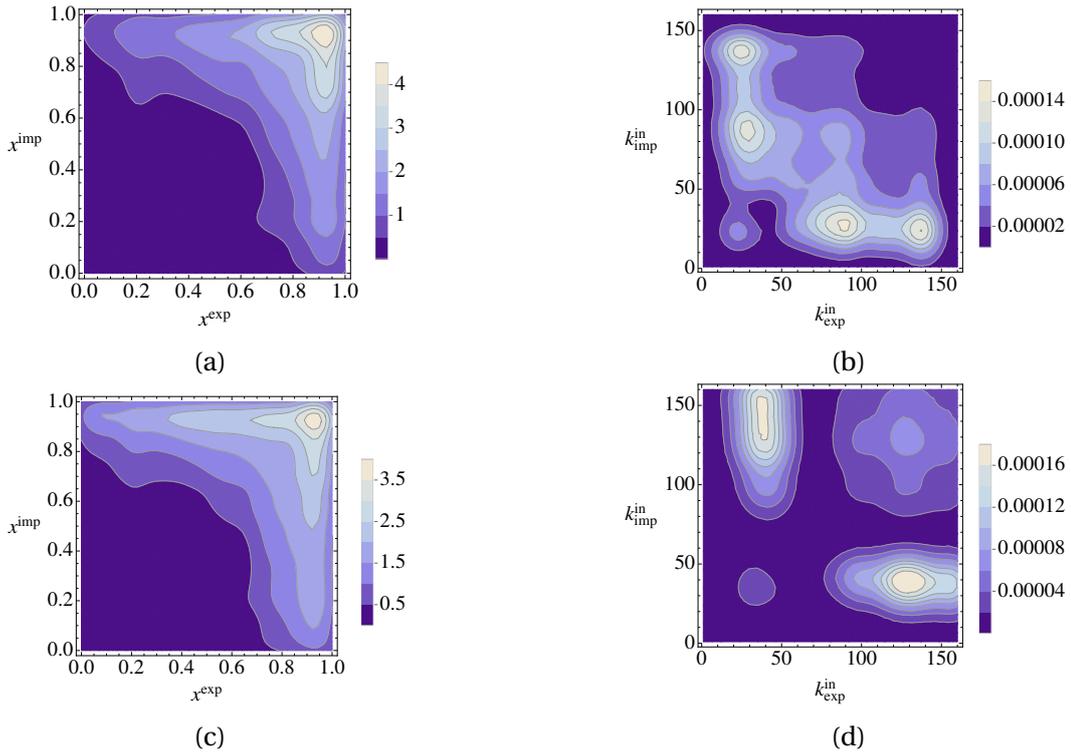
Figure 5.7: Contour plots of histograms of node properties adjacent to edges that are present in the years 1984 and 2000. (a) Fitness of nodes adjacent to edges in the year 1984. (b) In-degrees of nodes adjacent to edges in the year 1984. (c) Fitness of nodes adjacent to edges in the year 2000. (d) In-degrees of nodes adjacent to edges in the year 2000.

similar countries, low connected countries trade with highly connected ones and this is true for trade in both directions, the plots are almost symmetrical for import and export. Fig. 5.7 also shows that trade patterns have hardly changed between the years 1984 and 2000. Merely existing patterns have become more apparent. The fitness-fitness histograms in Figs. 5.7a and 5.7c show a slightly different pattern compared to the histogram in Fig. 5.6a. While edges are added most preferentially between countries of different fitness, the static view reveals that most of the trade is taking place between high fitness countries at the top of the ranking scale. Fig. 5.8 shows some example contours for different functional forms. It is clear that, if fitness is defined as in Eq. (5.13), the microscopic structure of trade in Fig. 5.7d cannot be explained with a multiplicative attachment kernel that has been suggested in [57] and is illustrated in Fig. 5.8c. The form of the fitness-fitness correlations in Fig. 5.7d is more similar to the functional forms in Figs. 5.8a and 5.8b.

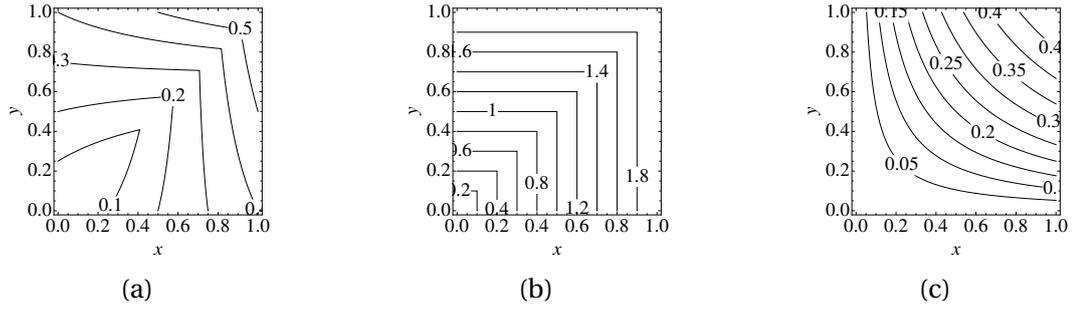The histogram of fitness in Fig. 5.7c can be used as an estimator for the true $C_{M,N}(x,y)$.

Figure 5.8: Contour plots of (a) $g(x,y) = 0.4|x-y| + 0.6xy$. (b) $g(x,y) = |x-y| + x + y$, (c) $g(x,y) = xy/(1+xy)$
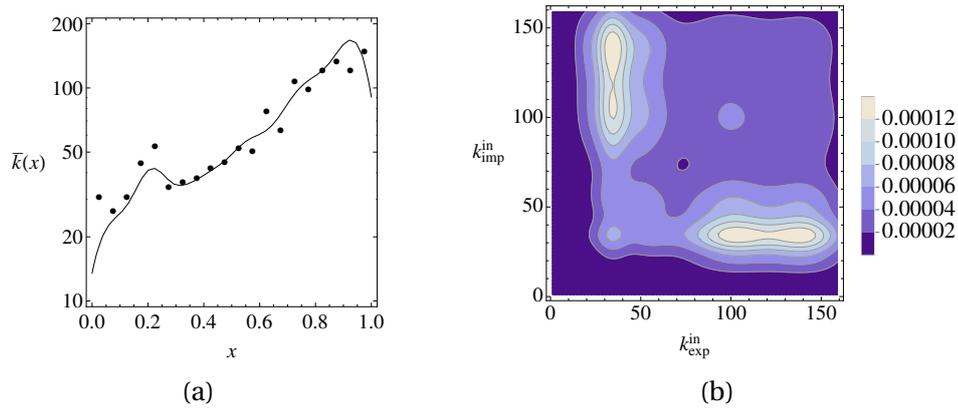

Figure 5.9: Theoretical results for the expected in-degree of a node and degree-degree correlations using empirical data as an estimator for the true $f(x,y)$. (a) Comparison of theoretical to analytical results of the fitness conditional expected in-degree of a randomly chosen node. The bullet points indicate the observed values. The solid line represents the estimator using the fitness-fitness histogram of nodes adjacent to edges in the year 2000 as an estimator for the true $C_{M,N}(x,y)$ that can be substituted in the expression for the average degree as in Eq. (5.12). (b) Contour plot of theoretical results on the degree-degree correlations $P(k,k')$, using Eq. (5.25) and the histogram of node's fitness adjacent to existing edges in the year 2000 as estimator for $C_{M,N}(x,y)$.

This edge density combined with the connection of $\lambda(x)$ and $C_{M,N}(x,y)$ in Eq. (5.7) can be utilised to estimate the resulting $\bar{k}(x)$ from the data and compare it with the actual $\bar{k}(x)$. This comparison can be found in Fig. 5.9a. Fig. 5.9 validates the hypothesis that the WTN is well described by a fitness model. The fitness conditional degree distribution, that is found in real data, is over a broad range of values confirmed by the values arising from the theory on static fitness models, using the empirical $\tilde{C}_{M,N}(x,y)$ as an estimator for $\lambda(x)$.

Not only on the aggregated level of the conditional expectation of in-degrees, good agreement of degree correlation structure of the network can also be found. The probability that there exists a link from a node with in-degree k toward a node with in-degree k' is

defined as

$$P(k,k') = \int_0^1 \int_0^1 p(k|x)C_{M,N}(x,x')p(k'|x')dx\,dx'. \tag{5.25}$$

Fig. 5.9b shows the $P(k,k')$ contour that can be computed using Eqs. (5.7), (5.25) and the observed edge density $\tilde{C}(x,y)$. The comparison of Figs. 5.7d and 5.9b shows an excellent agreement between the theory and the empirical findings. The areas of high trading intensities are almost congruent.

The empirical analysis, that has been laid out in this chapter confirms the general understanding that the network of international trade is governed by a vertex intrinsic fitness model. However, the investigation into the microscopic detail –the fitness correlation matrix – of the network revealed that the attachment kernel $f(x,y)$ is less trivial than, for example, assumed in [57, 58]. The topology in a fitness driven network is determined by two variables, the attachment kernel $f(x,y)$ and the fitness density $\rho(x)$. For the sake of robustness, fitness has been defined in such a way, that one degree of freedom is eliminated and the model depends solely on the form of $f(x,y)$.

The focus will now be turned to the attachment kernel, that has been proposed in [57, 58]. It has been shown so far that the WTN is well described using an attachment kernel that at least partially favours links between dissimilar countries when fitness is defined as in Eq. (5.13). This does not imply that the attachment kernel $f(x,y;a) = axy/(1+axy)$, which was proposed in [57, 58] together with fitness definitions ϑ' and ϑ'' is incorrect. Moreover, it has been shown that this kernel yields correct predictions for derived quantities, such as the average nearest neighbour-degree and the fitness conditional degree distribution [57, 58].

Fig. 5.10 shows empirical results of the microscopic structure of the WTN using the alternative fitness definition ϑ'', defined in Eq. (5.19). Fig. 5.10c illustrates that the results become very noisy for larger values of fitness. The fractions of the sample that are included in the three-dimensional histograms are shown in the cumulative distribution of fitness in Fig. 5.10b. The data for the first ∼85% is relatively noise free. Fig. 5.10a illustrates that the attachment kernel is symmetric in its arguments and favours connections between high-fitness nodes. The emerging pattern is very different from the one that
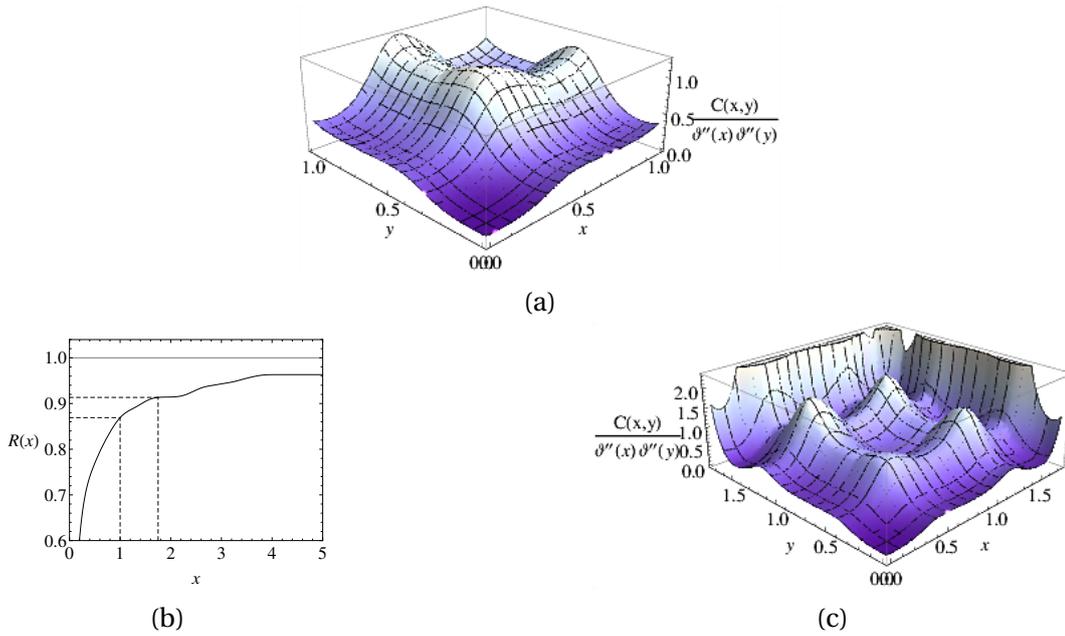
(a)

(b)

(c)

Figure 5.10: (a) Normalised connection density of the WTN in the year 2000 over the unit-square, using the fitness definition in Eq. (5.19). (b) Cumulative distribution function of fitness in the WTN in the year 2000 with fitness defined as in Eq. (5.19). The dashed lines mark the upper domain of the density plots in (a) and (c). (c) Normalised connection density of the WTN in the year 2000 over $[0,1.75]$, using the fitness definition in Eq. (5.19)

was found using fitness definition ϑ. Notably, $\tilde{C}_{M,N}(x,y)$ is non-zero at the boundaries $(x,0)$ and $(0,y)$. This is an indicator that $C_{M,N}(x,y)$ must be at least partially additive. Moreover, apart from being a technicality, a pure multiplicative kernel implies zero trade with with lowest fitness countries. This would represent a system in which trade advantages that arise from different income levels would not be monetarised. A more detailed analysis of the exact form of the edge density is not possible. Fig. 5.10c shows that the entire analysis with fitness defined as in Eq. (5.19) is not very robust. The necessary normalisation of the edge density $C_{M,N}(x,y)$ introduces a considerable amount of noise for larger fitness values. This shows that the introduction of fitness defined as in Eq. (5.13) is a significant advancement over previous definitions.

7 Conclusion

In this chapter, empirical aspects of international trade have been discussed. The microscopic structure of the world trade network, that is illustrated by computing the edge

density matrix, exhibits a pattern which is significantly different from previously assumed ones. It has been found that the income dependency of trade between two countries is not purely multiplicative, but also at least partially additive with a preference for trade between dissimilar countries. It has also been shown here that the static hidden variable model gives an excellent description of the world trade network on the microscopic level. So far, only the aggregate level had been studied. For future analysis, a novel approach to define fitness has been proposed here. By introducing a rank ordering of income, one degree of freedom, namely the fitness density, is eliminated. This reduces the noise in the statistical analysis significantly. Additionally, it has been demonstrated that empirical results on the world trade network depend largely on the chosen data source. Even first order results, such as the degree distribution can differ depending on the data provider. The magnitude of these disparities is larger than expected and should be considered in future investigations.

8   Countries Covered in this Study

Afghanistan	Costa Rica	Haiti	Mongolia	Slovenia
Albania	Cote Divoire	Honduras	Morocco	Somalia
Algeria	Croatia	Hungary	Mozambique	South Africa
Angola	Cuba	Iceland	Myanmar	Spain
Argentina	Cyprus	India	Nepal	Sri Lanka
Armenia	Czech Rep	Indonesia	Neth.Ant.Aru	St.Kt-Nev-An
Australia	Czechoslovak	Iran	Netherlands	Sudan
Austria	Dem.Rp.Congo	Iraq	New Calednia	Suriname
Azerbaijan	Denmark	Ireland	New Zealand	Sweden
Bahamas	Djibouti	Israel	Nicaragua	Switz.Liecht
Bahrain	Dominican Rp	Italy	Niger	Syria
Bangladesh	Ecuador	Jamaica	Nigeria	Taiwan
Barbados	Egypt	Japan	Norway	Tajikistan
Belarus	El Salvador	Jordan	Occ.Pal.Terr	Thailand
Belgium-Lux	Eq.Guinea	Kazakhstan	Oman	Togo
Belize	Estonia	Kenya	Pakistan	Trinidad Tbg
Benin	Ethiopia	Kiribati	Panama	Tunisia
Bermuda	Fiji	Korea D P Rp	Papua N.Guin	Turkey
Bolivia	Finland	Korea Rep.	Paraguay	Turkmenistan
Bosnia Herzg	Fm USSR	Kuwait	Peru	Uganda
Brazil	Fm Yemen AR	Kyrgyzstan	Philippines	UK
Bulgaria	Fm Yemen Dm	Lao P.Dem.R	Poland	Ukraine
Burkina Faso	Fm Yugoslav	Latvia	Portugal	Untd Arab Em
Burundi	France,Monac	Lebanon	Qatar	Uruguay
Cambodia	Gabon	Liberia	Rep Moldova	USA
Cameroon	Gambia	Libya	Romania	Uzbekistan
Canada	Georgia	Lithuania	Russian Fed	Venezuela
Cent.Afr.Rep	Germany	Madagascar	Rwanda	Viet Nam
Chad	Ghana	Malawi	Samoa	Yemen
Chile	Greece	Malaysia	Saudi Arabia	Yugoslavia
China	Greenland	Mali	Senegal	Zambia
China HK SAR	Guatemala	Malta	Seychelles	Zimbabwe
China MC SAR	Guinea	Mauritania	Sierra Leone	Fm Yemen Ar
Colombia	GuineaBissau	Mauritius	Singapore	
Congo	Guyana	Mexico	Slovakia	

Table 5.1: List of 173 countries that are included in the study. The naming is directly taken from the UN-data archive, which leads to multiple entries for single countries, that were used in different years.

Chapter 6

Mutual Selection in Time-Varying Networks

Time-varying networks play an important role in the investigation of the stochastic processes that occur on complex networks. The ability to formulate the development of the network topology on the same time scale as the evolution of the random process is important for a variety of applications including the spreading of diseases. Past contributions have investigated random processes on time-varying networks with a purely random attachment mechanism. The possibility of extending these findings towards a time-varying network that is driven by mutual attractiveness is explored in this chapter. Mutual attractiveness models are characterised by a linking function that describes the probability of the existence of an edge, which depends mutually on the attractiveness of the nodes on both ends of that edge. This class of attachment mechanisms has been considered before in the fitness based complex networks literature, but not on time-varying networks. Also, the impact of mutual selection is investigated alongside opinion formation and epidemic outbreaks. The voter model exhibits an unanticipated behaviour as the network never reaches consensus in the case of mutual selection, but stays forever in its initial macroscopic configuration, which is a further piece of evidence that time-varying networks are very different from their static counterpart with respect to random processes that take place on them. Another finding is that epidemic outbreaks are ac-

celerated by uncorrelated mutual selection compared to previously considered random attachment.

1   Introduction

Complex networks have been studied extensively over the last decade. However, this research is usually not conducted as an end in itself, but instead aims towards a better understanding of dynamical processes that take place on a network topology. Many contributions to these epidemiological models assume that the evolution of the network and the unfoldment of the dynamical process operate on two very different time scales, such that the network evolves more slowly than the dynamical process on top of it [13, 122]. However, this is a very restrictive assumption that does not hold in many circumstances. For a virus to spread, two individuals must be in contact when at least one of them is infected. The fact that these two individuals are linked in a static network, i.e. by friendship does not necessarily mean that the virus is passed on through their shared edge.

In recent years the field of temporal networks has attracted increasing attention and several different names have emerged in the literature, such as temporal graphs, dynamic networks, evolving graphs, time-varying graphs, . . . etc. [68]. The origin of these networks comes from real-world phenomena like face-to-face contacts or the phone call network, to name just a few. The central motivation is that edges are not persistent but can occur and be withdrawn within a short time. This has a direct impact on processes that take place on networks. While classical models consider the time-scale of the process to be much shorter than the time-scale of the network evolution [13, 97, 112, 113], the processes taking place on temporal networks have the same time-scale as fluctuations of the network topology. The issue that arises here becomes immediately evident if transitivity is taken into consideration. On a static graph edges are transitive, so if for example there exists a link (A, B), and a link (A, C), then there exists as well an indirect path between nodes B and C via node A. In a temporal network, this transitivity no longer exists if the edges (A, C) and (A, B) exist during different instances of the network [68].

A temporal network and its integration can be understood in terms of the daily social life of inhabitants of a town. Each day forms an instance of a temporal network, that starts initially empty. Contacts met during the day build the links in this instance. On the following day the network instance of social contacts starts empty again. All these network instances for many days can be joined to build the integrated network, which aggregates all instances to a social graph that describes the network of acquaintances. This integrated network is what is observed if one analyzes the social graph of Facebook. However, taking the single instances into account rather than the aggregation as it is done in the classical complex networks literature adds information as outlined above. In this particular example one could think of the spread of a rumor. Modeling the spread of a rumor with the integrated, static network will bias the outcome, since the rumor will not spread across edges that are not active at the time the rumor arrives at a particular node.

An excellent overview of the various applications of temporal networks in many disciplines can be found in [68]. The applications range from person-to-person communication [71, 96, 121], cell biology and brain networks [64, 80], to aspects of distributed computing and seasonal food webs [68]. Another example of a time-varying network is the evolution of groups. In- and outflux of groups in conferences have been studied in [117, 131].

One essential difference between temporal networks and their static counterpart is the development of node centrality. Node centrality measures the importance of a node within a graph. Measuring centrality is for example used when devising an optimal vaccination strategy or designing a successful marketing campaign. In classical networks, that are driven by preferential attachment for instance, the centrality of a node is monotone: new nodes connect preferentially to central nodes in the network. It is therefore correct to assume that once a central node is found it will stay central for the life-time of the network. Therefore this node is assumed to serve well as an entry point for all sorts of campaigns. Temporal networks are very different in this respect. A node that is central in one instance can be entirely disconnected in the following time step. Therefore successful immunization strategies or information campaigns will need to take these fluc-

tuations into account [67].

Recently, a model of a temporal fitness driven network has been investigated in [100]. Every node in that network is endowed with an intrinsic activity random variable that controls the probability that this node becomes active and initiates m new links in an instance of the network. The targets of these new links are chosen at random. However, several networks have been found to be driven by mutual agreement rather than just one sided selection, the world trade network that has been discussed in the previous chapter is one example [57]. An earlier example of a temporal network that is driven by mutual selection, encountering an adverse process in this topology is presented in [77]. The network model under consideration is the formation of sexual relationships and the aim of the study is to investigate the pattern of spreading diseases on this network. Pairs in this network are not chosen randomly, but with probability depending on a function $\phi(k_1, k_2)$ that depends on the node degree, hence on the number of already existing sexual relationships. Diffusion and transport problems on this type of network have been studied in [82, 100]

In this chapter, the work of Perra et al. [100] is extended by taking mutual preferences during the edge formation process into account. The effect of mutual selection on the topology is already well understood for static [24, 111] and dynamic networks (see chapter 3). As well as the degree distribution, epidemic spreading is studied here and compared to previous results in [100]. Finally, this chapter extends the body of existing literature on temporal network with a study of opinion formation. The strikingly surprising result is that consensus is unobtainable in a time varying network independent of the form of the attachment kernel. The analytical results for the degree distribution and epidemic spreading are derived for the special case of a separable attachment kernel. All analytical findings are backed by extensive numerical simulations.

2   Model

The network model consists of N nodes, that are endowed with an intrinsic random fitness x_i, drawn from a probability density $\rho(x)$. Every instance of the network $\mathscr{G}_t$ is initialised with these N nodes without edges and persists for a time span of Δt. With probability $a_i \Delta t$ a node i becomes active and originates m links, where a_i is the activity parameter of a node, defined as $a_i = \eta x_i$, and η is a normalization coefficient to fix the average number of active nodes per time step to $\eta N \langle x \rangle$. With probability $1 - a_i \Delta t$ a node remains inactive, but can receive links from other active nodes. The destination of a link depends mutually on the fitness of the origin of the link x_i and the fitness of the destination x_j, expressed in terms of an attachment kernel $f(x_i, x_j)$.

The integrated network G_T is the union of all network instances prior to time T: $G_T = \bigcup_{t=0}^{T} \mathscr{G}_t$. Multiple links, i.e. links that occurred during several instances are not counted several times. Merely the existence of a link in any given instance $\mathscr{G}_t$ affects the existence of a link in the integrated network.

2.1 Degree Distribution

Following the dynamics of all single instances of the network, the degree distribution of the integrated network at time T is of central interest. For the final degree distribution, edges that are repeated in various instances of the network are counted as one. It will be illustrated later in this text that the impact of these multiple edges is anyway negligible.

The probability that a node i initiates a link towards a node j during Δt is defined as

$$\Omega_{x_i \to x_j}^{\Delta t} = a_i m \Delta t \cdot \frac{f(x_i, x_j)}{N \int_0^\infty f(x_i, \xi) \rho(\xi) d\xi}. \tag{6.1}$$

That is the probability that node i becomes active multiplied with the probability that a node with fitness x_i initiates a link towards a node with fitness x_j. Further, denote the probability that there exists a link between nodes i and j in the aggregated network at

time t with $p_{ij}(t)$ and the probability of non-existence with $\bar{p}_{ij}(t)=1-p_{ij}(t)$. Since duplications of links during the aggregation process of all temporal instances are possible, it is straightforward to write down an equation for the probability that there exists no link between nodes i and j, hence the following calculation is based on $\bar{p}_{ij}(t)$, which obeys this rate equation

$$\bar{p}_{ij}(t+\Delta t)=\bar{p}_{ij}(t)\cdot(1-\Omega^{\Delta t}_{x_i\to x_j}). \tag{6.2}$$

Using the boundary condition $\bar{p}_{ij}(0)=1$ since all nodes are initially disconnected, one obtains

$$\bar{p}_{ij}(t)=\exp\left\{-\frac{a_i m t f(x_i,x_j)}{N\int_0^\infty f(x_i,\xi)\rho(\xi)d\xi}\right\}. \tag{6.3}$$

The aggregated network becomes a trivial fully connected graph for t/N close to 1. Assuming a sparse network, i.e. $t\ll N$ leads to

$$\bar{p}_{ij}(t)=1-\frac{a_i m t f(x_i,x_j)}{N\int_0^\infty f(x_i,\xi)\rho(\xi)d\xi}. \tag{6.4}$$

The expected out-degree of a node at time T is given by

$$k^{\mathrm{out}}_T(i)=\sum_{j=0}^N p_{ij}(T)=Tma_i. \tag{6.5}$$

Since the expected number of links that are initiated during a given time step by a node with activity a is $\Delta t\,am$, the expected number of edges that originate from a node after $T/\Delta t$ time steps is given by Tma_i. Hence multiple edges that occur during the aggregation process can be neglected in the leading order approximation.

The total degree of a node is the sum of the in-degree and the out-degree without duplications. In order to properly count, the in-degree here is similarly defined as in [100]. It is the number of edges pointing towards a node in the integrated network which have

never been reciprocated. This quantity can be expressed solely in terms of $p_{ij}(t)$

$$k_T^{\text{in}}(i) = \sum_{j=0}^{N} p_{ji}(T) \cdot \bar{p}_{ij}(T) \tag{6.6}$$

$$\simeq \frac{mT}{N} \sum_{j=0}^{N} \frac{a_j f(x_j, x_i)}{\int_0^\infty f(x_j, \xi)\rho(\xi)d\xi}. \tag{6.7}$$

The total degree of a node is then given by $k_T(i) = k_T^{\text{in}}(i) + k_T^{\text{out}}(i)$:

$$k_T(i) = Tma_i + \frac{mT}{N} \sum_{j=0}^{N} \frac{a_j f(x_j, x_i)}{\int_0^\infty f(x_j, \xi)\rho(\xi)d\xi}. \tag{6.8}$$

Rewriting this equation in a continuous fitness representation which is approximately correct for large N, one obtains

$$k(x) = Tm\eta x + Tm\eta \int_0^\infty \frac{\sigma f(\sigma, x)}{\int_0^\infty f(\sigma, \xi)\rho(\xi)d\xi}\rho(\sigma)d\sigma. \tag{6.9}$$

Assuming a factorisable form of $f(x,y) = g(x)h(y)$, Eq. (6.9) becomes

$$k(x) = Tm\eta \cdot \left(x + h(x) \cdot \frac{\langle x \rangle}{\langle h \rangle} \right) \tag{6.10}$$

$$\text{where } \langle h \rangle = \int_0^\infty h(x)\rho(x)dx. \tag{6.11}$$

The form of the degree distribution $p(k)$ depends on the choice of $f(x,y)$, respectively $g(x)$ and $h(y)$. Consider for example $f(x,y) = xy$. In this case, Eq. (6.10) becomes

$$k_T(i) = 2Tma_i. \tag{6.12}$$

For large enough T and N this relation can assumed to be monotonous, thus the inverse

$$x(k) = \frac{k}{2Tm\eta} \tag{6.13}$$

exists. This simplification has also been made and proven as useful for instance in [16,

24]. Using now that $p(k)dk \sim \rho(x)dx$, then

$$p(k) = \rho(x(k))\frac{dx}{dk} \tag{6.14}$$

$$= \rho\left(\frac{k}{2Tm\eta}\right) \cdot \frac{1}{2Tm\eta}. \tag{6.15}$$

This shows that the form of $p(k)$ is inherited directly from the form of $\rho(x)$. This result is in common with the static fitness model in [111] for the case $f(x,y) = xy$. Substituting $h(x) = 1$ into Eq. (6.10) replicates the result of [100]. In order to confirm this result, a numerical simulation was carried out, with $\rho(x) = e^{-x}$, $N = 50,000$, $T = 1,000$, $m = 2$ and $\eta = 0.1$. The theoretical results are in excellent agreement with the numerical results over a wide range of k as Fig. 6.1 shows.
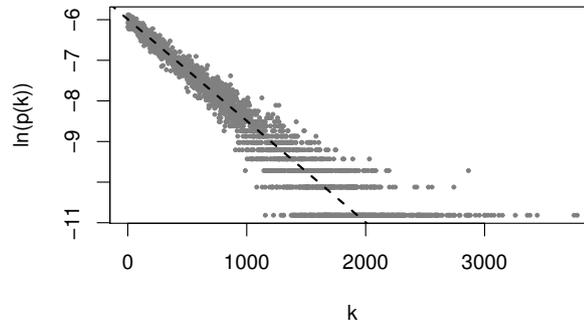

Figure 6.1: Results from numerical simulation with $f(x,y) = xy$, $\rho(x) = e^{-x}$, $N = 50,000$, $T = 1,000$, $m = 2$ and $\eta = 0.1$, compared to the prediction (dashed) given in Eq. (6.15).

2.2 Epidemic Spreading

This subsection is dedicated to the problem of epidemics on the network and discusses the SIS model with transition probability λ and recovery time $1/\mu$ [97]. The quantity of interest here is the number of infected individuals at time t, denoted by I^t. The dynamics of this quantity can be calculated by modeling the dynamics of the number of infected individuals with activity level a at time t, denoted with I_a^t.

The number of infected individuals of class a obeys the recurrence relation

$$
\begin{aligned}
I_a^{t+\Delta t} = {} & I_a^t - \mu \Delta t I_a^t \\
& + \lambda (N_a - I_a^t) \cdot m a \Delta t \cdot \sum_{a'} \frac{f(a,a')}{N \int_0^\infty f(a,\xi) \rho(\xi) d\xi} \cdot I_{a'}^t \\
& + \lambda (N_a - I_a^t) \cdot \sum_{a'} I_{a'}^t m a' \Delta t \frac{f(a',a)}{N \int_0^\infty f(a',\xi) \rho(\xi) d\xi}.
\end{aligned}
\tag{6.16}
$$

With N_a as the number of nodes in activity class a. The second term corresponds to the recovery of an infected individual. The third term quantifies the effect of a susceptible node to become active and connect to an infected individual. To be more precise there are $(N_a - I_a^t)$-many susceptible individuals with activity level a, each of these becomes active with probability $a\Delta t$. The term inside the sum is then the probability to connect to an individual of activity level a' that is infected. The fourth term accounts for the possibility that a susceptible node of activity level a becomes infected by receiving a link from an infected node. Hence the term inside the sum corresponds to infected nodes of activity level a', that connect to node with activity a and pass on the infection via this link with rate λ.

The total number of infected individuals can be obtained using

$$
I^t = N \int_0^\infty I_a^t \rho(a) da.
\tag{6.17}
$$

The epidemic threshold is a condition on λ and μ, that represents the tipping point between an endemic and the extinction of a virus on a network. Much research has been conducted, characterizing the epidemic threshold for particular network topologies [13]. Independent of the underlying topology of the network, the epidemic threshold is given by the highest eigenvalue of the adjacency matrix of the network [28, 123]. However, information on the adjacency matrix is not available in the present context of a mean-field approximation. A concept for the mean-field approach that follows elementary considerations of mathematical analysis has been illustrated for example in [13]. This approach will be used in the following as well.

In order to find a closed form expression for the epidemic threshold, a factorisable, not

necessarily symmetric form of f will be considered in the following

$$f(x,y) = g(x)h(y).$$

Behind that the following definitions will be used for brevity

$$\gamma^t = \sum_a h(a)I_a^t \quad \text{and} \quad \theta^t = \sum_a I_a^t a.$$

Using these conventions, Eq. (6.16) becomes

$$
\begin{aligned}
I_a^{t+\Delta t} = {} & I_a^t - \mu \Delta t I_a^t \\
& + \lambda \cdot \frac{1}{N}(N_a - I_a^t)m a \frac{\Delta t}{\langle h \rangle} \sum_{a'} h(a')I_{a'}^t \\
& + \lambda \cdot \frac{1}{N}(N_a - I_a^t)m \frac{h(a)}{\langle h \rangle} \Delta t \sum_{a'} a' I_{a'}^t.
\end{aligned}
\tag{6.18}
$$

The quadratic terms in I_a^t can be neglected in the leading order approximation, since the main quantity of interest is the onset of the epidemic, where I_a^t is sufficiently small. The epidemic threshold can then be obtained without further simplifications. Summing Eq. (6.18) over a and taking the continuous time limit, one obtains

$$\frac{\partial I}{\partial t} = -\mu I + \lambda \gamma^t m \frac{\langle a \rangle}{\langle h \rangle} + \lambda \theta^t m. \tag{6.19}$$

As a next step, two more equations are introduced by multiplying Eq. (6.18) with a and summing to obtain

$$\frac{\partial \theta}{\partial t} = -\mu \theta + \lambda \gamma^t m \frac{\langle a^2 \rangle}{\langle h \rangle} + \lambda \theta \frac{\langle a h \rangle}{\langle h \rangle} m. \tag{6.20}$$

By multiplying Eq. (6.18) with $h(a)$ one obtains

$$\frac{\partial \gamma}{\partial t} = -\mu \gamma + \lambda \gamma \frac{\langle a h \rangle}{\langle h \rangle} m + \lambda \frac{\langle h^2 \rangle}{\langle h \rangle} \theta^t m. \tag{6.21}$$

The resulting system of linear differential equations can be written as

$$
\begin{bmatrix} \frac{\partial I}{\partial t} \\ \frac{\partial \theta}{\partial t} \\ \frac{\partial \gamma}{\partial t} \end{bmatrix} = \begin{bmatrix} -\mu & \lambda m & \lambda \frac{\langle a \rangle}{\langle h \rangle} m \\ 0 & -\mu + \lambda \frac{\langle a h \rangle}{\langle h \rangle} m & \lambda \frac{\langle a^2 \rangle}{\langle h \rangle} m \\ 0 & \lambda \frac{\langle h^2 \rangle}{\langle h \rangle} m & -\mu + \lambda \frac{\langle a h \rangle}{\langle h \rangle} m \end{bmatrix} \begin{bmatrix} I \\ \theta \\ \gamma \end{bmatrix}.
\tag{6.22}
$$

The solution to this matrix differential equation can be stated as a polynomial of exponentials of the eigenvalues of that matrix. Hence the value of the largest eigenvalue dominates the development of the disease, and controls whether it dies out or becomes endemic. The epidemic threshold is therefore given by

$$
\Lambda_m = 0.
\tag{6.23}
$$

Whereby Λ_m is the largest eigenvalue of above matrix. The eigenvalues are given by

$$
\Lambda_1 = \lambda m \frac{\langle a h \rangle}{\langle h \rangle} - \sqrt{\frac{\langle h^2 \rangle}{\langle h \rangle^2} \langle a^2 \rangle \cdot m \lambda} - \mu
\tag{6.24}
$$

$$
\Lambda_2 = \lambda m \frac{\langle a h \rangle}{\langle h \rangle} + \sqrt{\frac{\langle h^2 \rangle}{\langle h \rangle^2} \langle a^2 \rangle \cdot m \lambda} - \mu
\tag{6.25}
$$

$$
\Lambda_3 = -\mu.
\tag{6.26}
$$

The largest eigenvalue is Λ_2, and hence the disease becomes endemic if

$$
\frac{\lambda}{\mu} > \frac{1}{m} \cdot \frac{\langle h \rangle}{\langle a h \rangle + \sqrt{\langle h^2 \rangle \langle a^2 \rangle}}.
\tag{6.27}
$$

Substituting $h(a) = 1$, Eq. (6.27) recovers the result of [100]. A numerical simulation shows as well the correctness of the above result, see Fig. 6.2. The reproductive number of the network can further be established in the same way as in [100], using that the expected number of edges per unit time is $mN \langle a \rangle$, leading to an average node degree of

$$
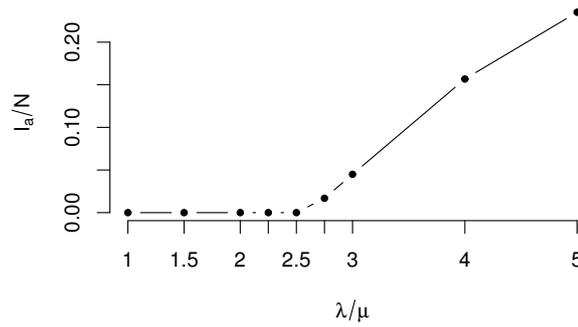\langle k \rangle_t = \frac{2 E_t}{N} = 2m \langle a \rangle.
\tag{6.28}
$$

Figure 6.2: Results from numerical simulation for the epidemic threshold. The plot shows the fraction of infected individuals I/N vs. λ/μ. The data is collected after $T = 10^4$ time steps in a network with $N = 10^4$ nodes, averaged over 100 simulations, with $\eta = 0.1$, $m = 1$, $\rho(x) = e^{-x}$ and $f(x,y) = xy$. The predicted epidemic threshold from Eq. (6.27) is $\lambda/\mu = 2.5$.

Then the critical reproductive number R_0^c is given by

$$R_0^c = \left[\frac{\lambda}{\mu}\right]^{\text{crit}} \cdot \langle k \rangle = \frac{2\langle a \rangle \langle h \rangle}{\langle a h \rangle + \sqrt{\langle h^2 \rangle \langle a^2 \rangle}}. \tag{6.29}$$

To investigate the effect that mutual selection has on the epidemic properties of the network, two cases will be considered in the following: $h(a) = 1$, that is the previously found result for random selection and $h(a) = a$, as the simplest form of mutuality in the selection process. The epidemic threshold for mutual selection is lower than for random selection if

$$\frac{\langle a \rangle}{2\langle a^2 \rangle} < \frac{1}{\langle a \rangle + \sqrt{\langle a^2 \rangle}}. \tag{6.30}$$

Assuming strictly positive fitness, the only solution to this inequality is given by

$$\langle a^2 \rangle > \langle a \rangle^2. \tag{6.31}$$

Hence the epidemic spreads faster in the case of mutual selection if this condition is fulfilled. Notice that Eq. (6.31) is a standard result of elementary probability theory. Thus epidemic spreading is always accelerated by mutual selection in the case of a factorisable attachment kernel.

2.3 Consensus Formation

Consensus formation is another commonly considered random process, that is usually investigated on a network-type topology. Unlike results on epidemic outbreaks that were studied in the previous subsection, results for voter models on time-varying networks do not exist in the complex networks literature. However, results for the voter model on heterogeneous graphs exist [88, 112, 113], which will serve as a benchmark. The quantity of interest in this section is the number of individuals in the network with fitness x and positive opinion at a given time step. $N_+(x, t)$, and the traditional voter model [13] will be considered. The number of individuals with negative opinion and fitness x will be denoted with $N_-(x, t)$.

In the traditional voter model, at every time step one node is chosen randomly and adopts the opinion of a randomly chosen neighbour. If the chosen node is not active and has not received any links during a network instance, its opinion remains the same. The number of positive opinions among nodes with fitness x evolves as follows

$$
\begin{aligned}
N_+(x, t + \Delta t) = {} & N_+(x, t) \\
& + \rho(x) \frac{N_-(x, t)}{N(x)} \sum_{j=1}^{N} (\Omega_{x_j \to x}^{\Delta t} + \Omega_{x \to x_j}^{\Delta t}) \frac{N_+(x_j, t)}{N(x_j)} \\
& - \rho(x) \frac{N_+(x, t)}{N(x)} \sum_{j=1}^{N} (\Omega_{x_j \to x}^{\Delta t} + \Omega_{x \to x_j}^{\Delta t}) \frac{N_-(x_j, t)}{N(x_j)}.
\end{aligned}
\tag{6.32}
$$

The second term on the right-hand side of Eq. (6.32) accounts for the increase in the number of positive opinions in the class of nodes with fitness x. The first part of that term accounts for the probability that a randomly chosen node has fitness x and negative opinion, the remainder of the term inside the sum represents the expected fraction of positive nodes in the neighbourhood of x, which is equal to the probability of choosing a neighbour with positive opinion, since opinions are dichotomous. The third term of Eq. (6.32) represents the opposite process to that in the second term.

One aspect of voting consensus in temporal networks can already be inferred from Eq. (6.32). The consensus, if at all, is reached very slowly. The terms inside the sums of Eq. (6.32)
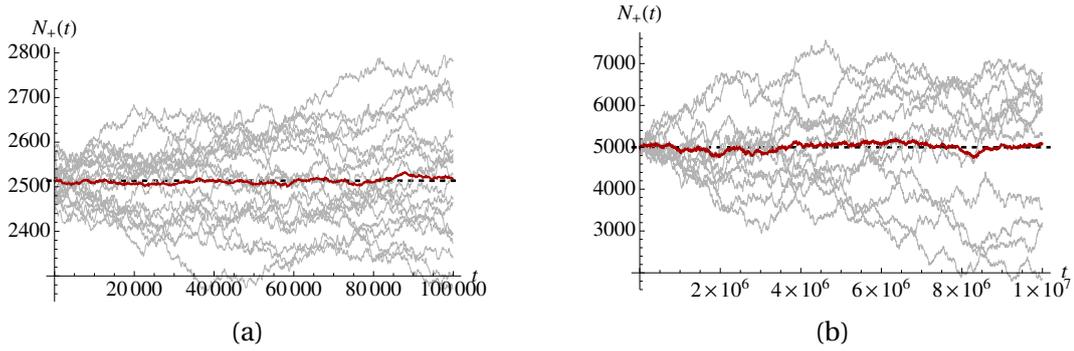
Figure 6.3: Traces of $N_+(t)$ for different configurations. Individual realisations in grey, average in red, initial condition is marked with a dashed line. $f(x,y) = xy$, $\rho(x) = e^{-x}$ and $\eta = 1/10$ for both cases. (a) 20 independent simulations with $m = 6$, $T = 10^5$ and $N = 10^4$. (b) 12 independent simulations with $m = 5$, $T = 10^7$ and $N = 10^4$.

are of order $1/N^2$, since $\Omega_{x_i \to x_j}^{\Delta t}$ is of order $1/N$ and $N(x_j) = N\rho(x_j)$. Hence the change of $N_+(x,t)$ is of order $1/N^3$. Define now the probability that two nodes with fitness x and x_j are linked as

$$\Psi(x, x_j) = \Omega_{x \to x_j}^{\Delta t} + \Omega_{x_j \to x}^{\Delta t} \tag{6.33}$$

and use $N_+(x,t) + N_-(x,t) = N\rho(x)$, therefore Eq. (6.32) can be rewritten to

$$\begin{aligned}
\frac{\partial N_+(x,t)}{\partial t} &= \rho(x) \sum_{j=1}^{N} \Psi(x, x_j) \frac{N_+(x_j, t)}{N\rho(x_j)} \\
&\quad - \frac{N_+(x,t)}{N} \sum_{j=1}^{N} \Psi(x, x_j).
\end{aligned} \tag{6.34}$$

Initially, the network is perfectly mixed and the number of nodes with positive opinion and fitness x is proportional to the number of nodes with fitness x. Therefore, the initial condition is given by

$$N_+(x, 0) = aN\rho(x), \quad 0 < a < 1. \tag{6.35}$$

Whereby the value of a depends on the specific configuration. Substituting this condition into Eq. (6.34) leads to

$$\frac{\partial N_+(x,t)}{\partial t} = 0. \tag{6.36}$$

This means that consensus can never be reached, moreover, the system stays forever in its initial configuration. Numerical simulations have been conducted to verify this result, the development of $N_+(t)$ for two different configurations can be found in Fig.(6.3). The

fact that consensus can never be reached on a time-varying network shows very clearly how different this class of networks is from classical networks. Sood, Antal, and Redner [112] have shown that the time until consensus is reached T_N, starting from an equally mixed population on a complete graph is $N \ln 2$ and for heterogeneous networks with degree distribution $p(k) \propto k^{-\alpha}$ with $\alpha < 3$, $T_N \ll N$. Thus consensus is usually reached reasonably fast. However, it could already been inferred from Eq. (6.32), that due to time-varying nature, the opinion update process is slower by several orders of magnitude compared to classic networks. That consensus is unobtainable is a surprising result.

3   Discussion

The field of time-varying networks is still in its early developments, but is certainly of great importance for many real-world applications. The ability to formulate network evolution and a topology coupled random process on the same time-scale enhances the understanding of many real-world phenomena. In this chapter, the concept of mutual selection on time-varying networks has been introduced. Mutual selection is a concept that is already understood well for static and dynamic network models, but has not been considered on time-varying networks previously. This study has shown that mutual selection has a direct impact on the way, epidemics spread on the network. It has also been pointed out that independent of the attachment kernel, consensus on a dichotomous decision can not be reached within the framework of a time-varying topology. Studying more involved random processes is left for future work.

Chapter 7

Cluster Formation Dynamics Based on Fitness

Unlike in the previous chapters, a dynamic model of cluster formation without underlying geometry is introduced in this chapter. Agents decide to join or leave clusters depending on their relative hierarchical position within a cluster. The model is motivated by an analogy to firm creation. The whole system receives a fixed amount of reward. The reward share of individual clusters depends only on their size. Each agent is endowed with a static random variable that describes some intrinsic quality. Outside a cluster, an agent cannot exploit this quality because of the continuum of necessary comparisons. Inside a finitely sized cluster, quality determines the hierarchical position of an agent. A higher hierarchical position translates into a higher reward. Depending on a comparison between outside and inside reward shares, each agent makes a decision to stay within a cluster or to leave. Within the bounds of this model, the impact of mobility imperfections is studied and it is found that the evolution of the model becomes increasingly predictable, the stronger the imperfections become. The size of the largest cluster in the system is smaller than the system by several orders of magnitude. The chapter is concluded with insights into a possible mechanism that can relax this constraint.

96

1   Introduction

The study of cluster formation and herding has been used to explain various phenomena, such as the sharing of information in financial markets [38, 45, 104], the distribution of firm sizes [39, 40], and dynamics of conflicts [119]. Herding into clusters is the phenomenon of individuals grouping together for a common goal. In the case of financial markets, traders stand together to share information. In the case of conflicts, herds emerge to do a common strike. Also customers can herd into groups of clients at one particular company, as discussed in [40]. In general, clusters grow if there exists a benefit in doing so. A large army can strike harder than a small one, and a trader that learns about information from others is certainly more likely to succeed than an isolated one. However, once a goal is reached, a herd can fragment, if there is no additional benefit for agents to remain inside a component. This fragmentation can occur either suddenly [38, 45, 104, 119] or gradually. In this chapter, a gradual fragmentation is discussed. Rather than assuming that remaining inside a cluster is not beneficial for each participant, it is assumed here, that the benefit differs among the agents inside the cluster. Another difference is the underlying structure. Previous models assume a network structure at the microscopic level. Here, agents compare each other with everybody else inside a component. Therefore, clusters would form complete graphs. This completeness does not add any additional information to the process, hence it will be abstracted from it entirely and no underlying geometry will be assumed.

The rules of evolution inside the system are motivated by the dynamics in an economy that might explain the origin of firms. It is a long debated question in economics why firms exist. This is a valid question, since firms are composed of individuals and it is not clear from the outset, why people organise themselves into firms, instead of remaining sole proprietors and contract their work with others. Classical approaches in the economics literature are based on problems of coordination between different agents, that include problems of work incentives, profit sharing and related issues [6, 33, 128]. For an in-depth review, see for instance [114].

Unlike the approaches based in economics, here a discussion of the problem that is motivated by microscopic dynamics of cluster formation is provided, which is based on actions of individual agents. A static number of agents is assumed to face continuously the decision whether to join or to leave a cluster, that they currently inhabit. Each cluster is a collection of agents that may represent, in this context, one single firm. Each agent is characterised by an attribute, that will be called fitness. This fitness describes some intrinsic quality of the agent, for instance education. The size of the system is fixed to a number of N agents.

Inside a finitely sized component, fitness determines the hierarchical position of an agent. Each cluster of size k, receives a share of k/N of the system's reward. The fittest agent inside a cluster receives the largest share of the cluster's reward, while the least fit agent receives the lowest share respectively. Fitness can be exploited as a measure of hierarchy within the bounds of one cluster, because only a finite number of one-to-one comparisons is necessary. Outside clusters such a comparison does not make sense. Firstly, N is assumed to be a large number, hence conducting one-to-one comparisons of each possible pair is prohibitively extensive. Secondly, a comparison between isolated nodes and clusters of size larger than one cannot be well defined. The hierarchy inside clusters creates incentives for agents to join or to leave existing clusters, depending on a comparison between the possible rewards. It should be noticed that problems of these kind do not only emerge in the context of firm evolution, but the organisation of message boards on the internet might also be understood in this way. Fitness in that context describes the ability to write interesting comments. If there is a race for attention, then nodes with interesting stories will stay atop of the hierarchy, while other agents might search for new places to publish their commentary.

The remainder of this chapter is organised as follows: in section 2, a detailed description of the model is given. In the first subsection of section 3, a simplified model is discussed to elucidate the behaviour of the system when most of its complexity is stripped away. In the second subsection, results from numerical simulation for the full model are presented. Finally, in section 4, the chapter is closed with a summary and concluding remarks.

2   Model

Before turning to the detailed model description, Tab. 7.1 summarises the notation, that

is used hereafter. The present model comprises a static set of N agents. Each agent is

Symbol	Meaning
N	Number of agents inside the system
q_i	Intrinsic quality of an agent with label i
x_i	Fitness of an agent with label i
m_i	Rank of agent with label i inside a component
$r_k(m)$	Reward of an agent with rank m inside a component of size k
$N_k(t)$	Number of clusters with size k
$M(t)$	Number of clusters in the system
$k_{\max}(t)$	Size of the largest component at time t.
$H(x)$	Standard Heaviside function
β	Probability that an agent, which should make a move, eventually moves

Table 7.1: Notation

endowed with a static attribute. This attribute is a random variable and is a proxy of the

agents quality. How to measure this intrinsic quality is not clear. Therefore, in the present

model only a transitive ordering relation is assumed to exist. Whenever faced with two

agents, it is possible to decide which one has a better quality, and therefore higher fitness.

Transitivity assures that the ranking has no loops, hence an objective order for all agents

can be found following pairwise comparisons. Formally fitness is then defined as follows:

if q_i describes the intrinsic quality of an agent, then the fitness of this agent is given by

$$x_i = \frac{1}{N} \sum_{j=1}^{N} H(q_i - q_j) \tag{7.1}$$

whereby $H(y)$ is the Heaviside function. This concept follows the notion that has already

been illustrated in chapters 3 and 5. Using standard methods of probability, it follows

directly from this definition, that the distribution of fitness is standard uniform. This is

not restricting generality, but furthermore allows to understand fitness in a more generic

way, since only a transitive ranking is assumed to exist. This is more robust than assum-

ing a specific form of quality distribution, which might be difficult to measure.

Agents interact with each other within the bounds of a closed system. The system pro-

duces some exogenously determined reward, which is shared between the agents. Each agent can either be isolated or part of a component, together with other agents. Each component of size k receives a profit share of k/N, thus an isolated agent receives $1/N$. Ignoring the distribution of profits within a cluster, for a moment, the average profit per agent is independent of the size of the component, that it inhabits. This notion is realistic as it can be found for instance in empirical data of firm sizes. Fig. 7.1 illustrates the average pay per worker for the entire US economy and for one specific industry grouped by different firm sizes.
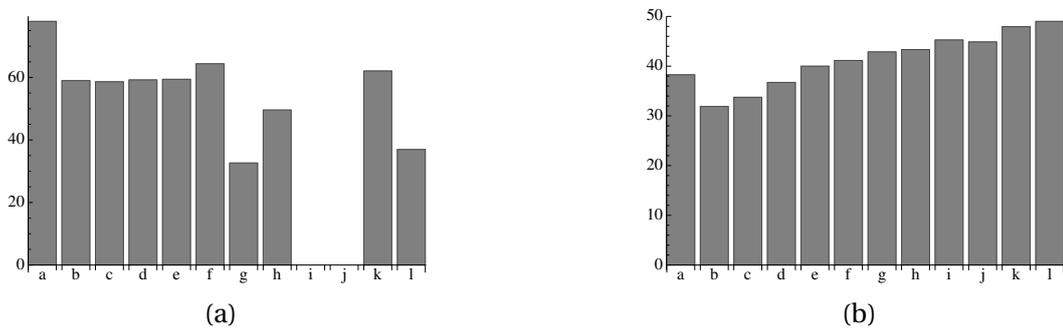

(a)   (b)

Figure 7.1: Average pay per worker for different aggregation levels. (a) shows the data for Industry 'Markerting Consulting Services' (NAICS:541613). Categories i and j are censored. (b) Shows data for all companies, aggregated over different industries. Data is taken from the US Census (2007). The company size categories on the x-axis are: a: $1 \leq k < 5$, b: $5 \leq k < 10$,c: $10 \leq k < 20$,d: $20 \leq k < 100$, e: $100 \leq k < 500$, f: $500 \leq k < 750$, g: $750 \leq k < 1000$, h: $1000 \leq k < 1500$, i: $1500 \leq k < 2000$, j: $2000 \leq k < 2500$, k: $2500 \leq k < 5000$, l: $5000 \leq k < 10000$.

The evolution of the system is then as follows: initially all agents are isolated. At each discrete time step, isolated agents can chose to join existing clusters, and agents inside of clusters can decide to leave. This decision is based on a comparison between inside and outside rewards. If the agent decides to remain a singleton, it faces the competition of the continuum and can therefore not exploit its intrinsic quality, thus its reward is just $1/N$. If an agent is inside a finitely sized component, its fitness translates into a ranking and its reward $r_k(m)$is proportional to its rank m. To be more precise, an agent with the m-th lowest fitness inside a component of size k receives a share of the firm's output in the amount of $r_k(m) = m/\sum_{j=1}^{k} j = 2m/[k(k+1)]$. This rule derives directly from postulating that $r_k(m) \propto m$ and that the sum of rewards inside a component is constraint to $\sum_{m'=1}^{k} r_k(m') = k/N$. Notice that ranking is increasing in fitness, such that $x_{(m)} < x_{(m+1)}$, where x_j is the $j-th$ highest fitness inside a component.

Each agent only joins a component, respectively stays inside a component if its reward is higher than if it were outside in a singleton component. The condition for joining, respectively staying inside a component of size k for an agent with label i is therefore

$$\frac{2m_i}{N(k+1)} \geq \frac{1}{N} \Leftrightarrow m_i \geq \frac{k+1}{2}. \tag{7.2}$$

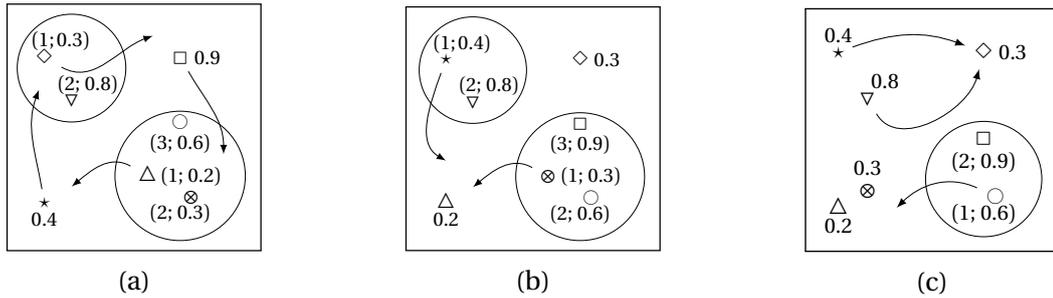Fig. 7.2 illustrates these mechanisms schematically.

Figure 7.2: Schematic illustration of the cluster evolution. The big circles represent clusters. Agents outside these clusters are annotated with their fitness. Inside clusters, the label contains additionally their rank. (a) Isolated agents decide to join clusters based on their prospective rankings. The agent with label $\star$, would rank in the upper left cluster as the second highest, and has therefore incentive to join. At the same time, the agent with label $\diamond$ has an incentive to leave because he ranks 1 in a cluster of size 2. The agent with label $\square$ has incentive to join the large cluster, because it would rank top in that cluster. At the same time, the agent with the lowest ranking decides to leave. (b) None of the isolated agent decides to join either cluster. But due to the maximisation principle of the agents, the agent with label $\star$ and the one with label $\otimes$ leave their cluster, because their rank is the below the threshold in their respective hierarchies. (c) The isolated agents with label $\star$, $\triangledown$ and $\diamond$ form a new cluster in the next time step. The agent with label $\bigcirc$ leaves her current cluster because of her relative hierarchical position. All other agents remain inactive in this instance.

Because of the reward sharing mechanism that assigns a higher share to larger clusters, agents prefer to join large clusters over small ones, because the possible payout is ex ante higher than in smaller clusters. Thus the probability that an isolated agent choses a cluster of size k in a given time step is kN_k/N, where N_k is the number of clusters of size k. At the end of each time step, choices are reviewed and agents withdraw from clusters if their profit share does not suffice. To be able to analyse also the outcome of mobility imperfections, a probability β is introduced that specifies the probability that an agent that is prone to leave a cluster according to the rules, eventually moves. A detailed, more formal description can be found in Algorithm 1.

```
 1  for t ← 1 to T do
 2  │   for agent ∈ {1, 2, ⋯ , N} do
 3  │   │   if isIsolated (agent) then
 4  │   │   │   cluster ← chooseRandomCluster ();
 5  │   │   │   m ← getRankOfAgent (agent, cluster);
 6  │   │   │   if m ≥ (|cluster| + 2)/2 then
 7  │   │   │   │   agent joins this cluster
 8  │   │   │   end
 9  │   │   end
10  │   end
11  │   for agent ∈ {1, 2, ⋯ , N} do
12  │   │   if ¬isIsolated (agent) then
13  │   │   │   cluster ← getClusterOfAgent (agent);
14  │   │   │   m ← getRankOfAgent(agent, cluster);
15  │   │   │   if m < (|cluster| + 1)/2 then
16  │   │   │   │   with probability β: agent leaves this cluster
17  │   │   │   end
18  │   │   end
19  │   end
20  end
```

Algorithm 1: Algorithmic description of the dynamical system. Each time step contains an intermediate state in which agents form together into clusters that are ex ante beneficial. After this, agents evaluate their ranks within these newly formed clusters and decide whether to stay or to leave the cluster again within this time step.

3   Results

3.1   Results for a Null Model without Fitness

Prior to the discussion of the results of the full model, a null model for later comparison will be briefly introduced. Consider a model with discrete time steps Δt without fitness. The model is investigated in the $\Delta t \to 0$ limit, so that just one action per time step can occur. Hence only one agent can become active. If this agent is inside a cluster of size greater than one, it will leave with probability β, if the agent is isolated, then it will join a cluster preferentially with size. The evolution of this model can be described with a set of two rate equations:

$$\frac{dN_1}{dt} = \left(1 - \frac{N_1(t)}{N} - \frac{2N_2(t)}{N}\right)\beta + 2\frac{2N_2(t)}{N}\beta - \frac{N_1(t)}{N}\left(1 - \frac{N_1(t)}{N}\right) - 2\frac{N_1(t)}{N}\frac{N_1(t)}{N} \quad (7.3)$$

$$\frac{dN_k}{dt} = -\beta\frac{kN_k(t)}{N} + \beta\frac{(k+1)N_{k+1}(t)}{N} + \frac{N_1(t)}{N}\frac{(k-1)N_{k-1}(t)}{N} - \frac{N_1(t)}{N}\frac{kN_k(t)}{N}, \quad k > 1$$

$$(7.4)$$

whereby $N_k(t)$ is the number of clusters of size k at time t. The first term of Eq. (7.3) corresponds to event, when an agent from a cluster of size $k \geq 3$ is chosen and leaves. The second term accounts for the increase in the number of isolated agents when a node inside a cluster of size $k = 2$ is chosen. The third term accounts for choosing an isolated agent and connecting it somewhere, not to another cluster of size $k = 1$. The last term represents the decrease of the number of isolated agents when an isolated agent is chosen, which connects then to another isolated one. The structure of Eq. (7.4) is similar. The first two terms account for a withdrawal of an agent from a cluster of size k, respectively $k + 1$. The third and fourth term account for the case, when an isolated agent is chosen and connects either to a cluster of size k or $k + 1$.

The non-linear system of differential equations in Eqs.(7.3-7.4) cannot be solved in closed form. The nonlinear terms cannot be neglected because of the large number of isolated agents inside the system. Thus, also the full model with fitness cannot be investigated purely analytically. However, this simplified model provides already an insight into some

of the macroscopic features that can also be found in the model with fitness. Fig. 7.3 shows results for the size of the largest cluster, the number of isolated nodes, as well as the number of distinct clusters and the cluster size distribution. It is clear from these
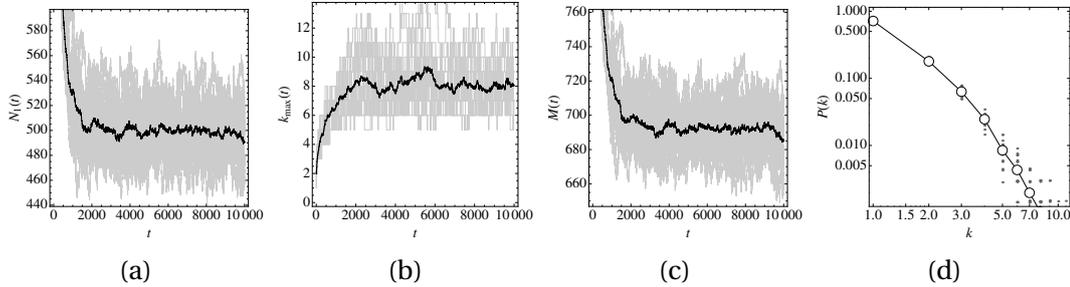

Figure 7.3: Simulation results for the null model. (a) Number of isolated nodes. (b) Size of the largest clusters. (c) Number of clusters. (d) Cluster size distribution. Black lines show average values after 20 independent simulations, the grey lines illustrate individual realisations. The size distribution in (d) is measured after 10^4 time steps in a system with 10^3 agents, and averaged over 20 realisations.

results that the model reaches a stationary state after an initial burn-in phase. As already mentioned above, Fig. 7.3a shows, that the number of isolated nodes is of similar order as the number of agents. Another feature that is evident, is the compressed domain of cluster size. The size of the largest cluster is around 7, while the size of system is $N = 10^3$.

3.2   Results of Fitness Based Cluster Dynamics

In this section, the numerical investigation of the full model with fitness is presented. Three major settings are under investigation here:

perfect mobility: $\beta = 1$, the system instantaneously corrects for any disfavoured states. If an agent has to move according to the rules, it will directly do so.

semi mobility: $\beta = 1/2$, agents only leave with 50% probability, when they are chosen to move according to the rules.

weak immobility: $\beta = 1/20$, agents are very immobile and leave a disfavoured state only with 5% probability.

It is shown later in the text, that immobility inside existing clusters implies immobility on the outside. This is due to the lack of available free agents that can possibly join. A further parameter for joining immobility would therefore be redundant.

Fig. 7.4 shows the development and the power spectrum of the time-series that describes the size of the largest cluster. The power spectrum decays at high frequencies as

$$S(\nu) \propto \nu^{-a} \tag{7.5}$$

With $a \cong 2.08$ for *perfect mobility*, $a \cong 0.95$ for the *semi mobility* regime and $a \cong 0.01$ for the *weak mobility* setting. Curiously, the power spectrum resembles almost white noise for the case of perfect mobility of agents. The stronger the mobility constraints are, the stronger the correlations also become in the system. If however, constraints are extremely high, i.e. in the case of $\beta = 1/20$, the short term correlations vanish, because of the lower frequency of actions inside the system. The extreme volatile behaviour, which becomes especially evident in the *weak mobility* setting is a direct consequence of the rules. Whenever the system is in a state of low $k_{\max}(t)$, then many agents are isolated and try to join existing clusters. This continues until either most of the isolated agents are absorbed, or only relative low fit agents are left outside. At this stage, the largest cluster is large, but the system is in a disfavoured state. Half of the agents inside this cluster are prone to leave, which is a consequence of Eq. (7.2). But in the case of $\beta = 1/20$, the probability that the size of the largest cluster can revert to its mean is very low. The system can correct itself only very slowly, hence the decline of $k_{\max}(t)$ occurs gradually. Once enough agents with sufficient high fitness are isolated, the build up process repeats itself, which leads to the volatile pattern that is observable. The case with the deterministic rule (*perfect mobility*) is different. The system can always come back to its desired configuration. Deviations from the mean are instantaneously corrected. This is illustrated in Fig. 7.4a.

Generally, the time series of $k_{\max}(t)$ tends to be constant over short intervals. These periods of motionlessness become stronger for smaller β. This phenomenon is a direct result from the rules, additions to the largest cluster occur relatively rarely. However, these periods might introduce spurious correlations that are not explanatory for the direction of the system. Similar to queuing theory, arrival numbers and times might be investigated

separately. To do so, a compressed sequence $k_{\max}^c(t)$ is introduced, which is defined by

$$k_{\max}^c = \left(k_{\max}(t') \colon k_{\max}(t') \neq k_{\max}(t'-1) \right)_{1 \leq t' \leq t}. \tag{7.6}$$

Fig. 7.4e illustrates the power spectrum of the compressed sequence $k_{\max}^c(t)$. The decay at high frequencies can - as in the case of the full series - be described by a power-law, only the exponents have changed. One aspect that has markedly changed is the sign of the exponent in the *perfect mobility* setting. Without compressing, the corresponding time series of $k_{\max}(t)$ does not show any structure. This is also confirmed by Fig. 7.5. Fig. 7.5 shows the partial autocorrelation function (PACF) of the two sequences $k_{\max}(t)$
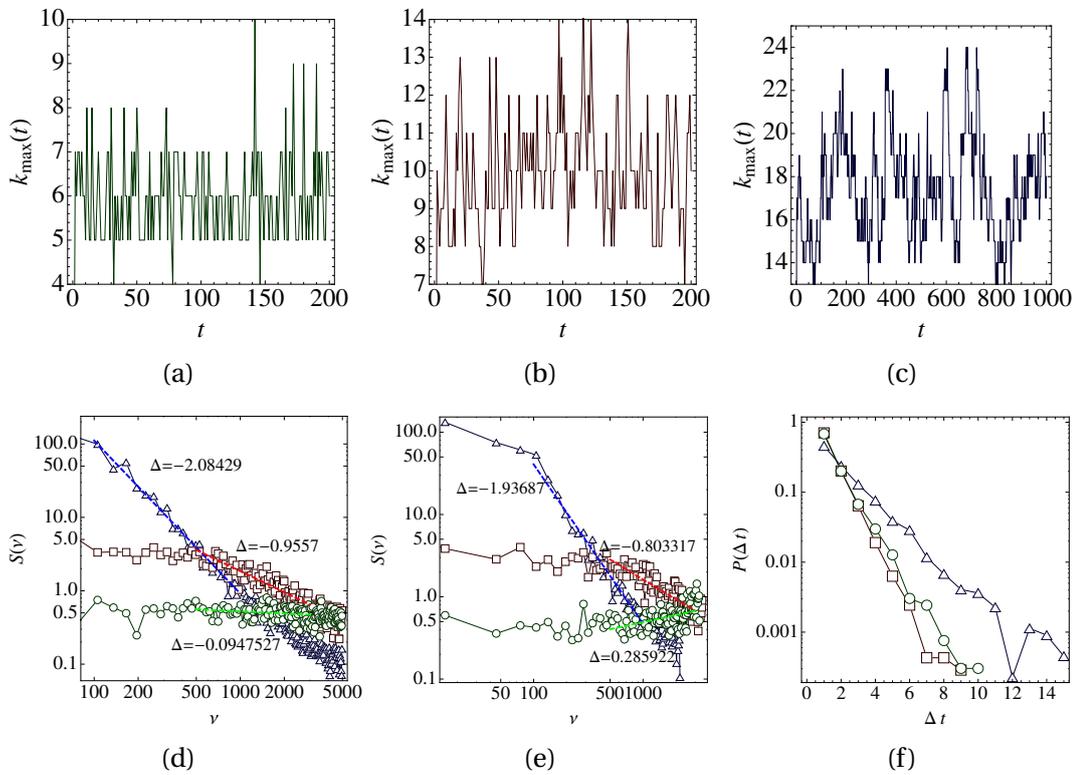

Figure 7.4: Time series results on the size of the largest cluster in a system with $N = 5,000$ agents. Development of the largest cluster for the (a) *perfect mobility* setting, (b) *semi mobility* setting and (c) the *weak mobility* setting. (d) Power spectrum of $k_{\max}(t)$ for the different configurations: *perfect mobility* ($\bigcirc$), *semi mobility* ($\square$), and *weak mobility* ($\triangle$). (e) Power spectrum of the compressed series of $k_{\max}^c(t)$ for the different configurations (f) Distribution of duration of stationarity of $k_{\max}(t)$ for the different configurations.

and $k_{\max}^c(t)$. The PACF is useful to identify autoregressive (AR) processes, which have the form

$$X_t = \sum_{\tau=1}^{p} \theta_\tau X_{t-\tau} + \epsilon_t \tag{7.7}$$

where p is the parameter of an AR(p) process. The PACF plot shows that an AR(1) process is describing the correlation structure well. Curiously, the correlations become negative for the *perfect mobility* setting. The difference between the correlation structure of the original sequence and the compressed one shows the usefulness of this transformation. The whole process is then characterised with the additional information of the durations of halting times. This distribution can be found in Fig. 7.4f. The domain is very short, so that reliable fit cannot be made. However, the decay is slightly heavier than exponential, but does not scale with a power.

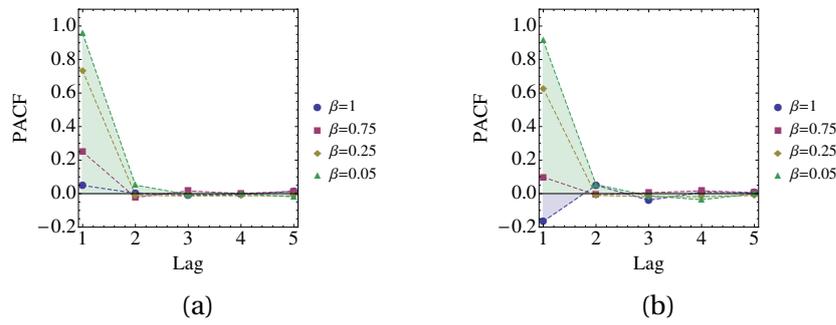(a)   (b)

Figure 7.5: PACF for (a) $k_{\max}(t)$ and (b) the compressed $k_{\max}^c(t)$ in a system with $N = 5000$ after $T = 10,000$ time steps for different values of β.

Although the time series of $k_{\max}(t)$ and $M(t) = \sum_k N_k(t)$ are stationary over time, their expectation depends on the total size of the system. The relationship between these quantities and the size of the system is illustrated in Fig. 7.6. Fig. 7.6 illustrates the impact

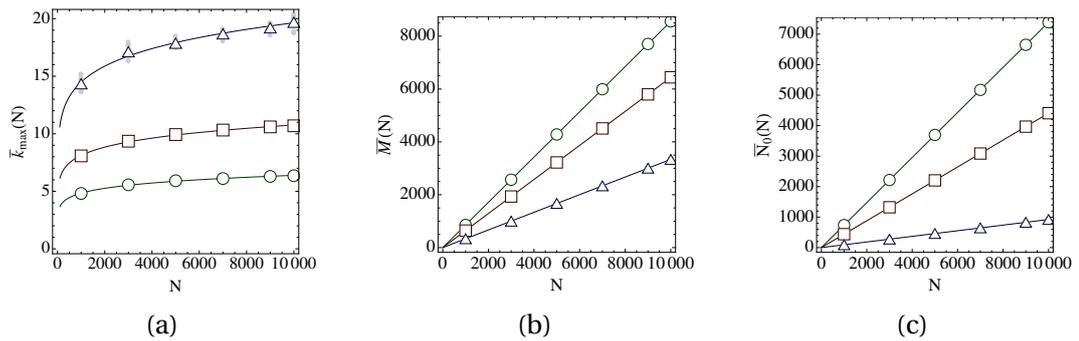(a)   (b)   (c)

Figure 7.6: (a) Average size of the largest cluster, (b) the number of distinct clusters and (c) the number of isolated agents for different system sizes after 10,000 time steps. The different settings are labelled as follows: *perfect mobility*: $\bigcirc$, *semi mobility*: $\square$, *weak mobility*: $\triangle$. The curves $\bar{k}_{\max}(N) = a N^b$ have parameters $(a, b) = (2.104, 0.121)$ for the *perfect mobility* setting, $(a, b) = (3.49, 0.122)$ for *semi mobility* and $(a, b) = (5.81, 0.132)$ for the *weak mobility* setting. The average number of clusters can be described by a linear fit $\bar{M}(N) = c + dN$, with $(c, d) = (0.108, 0.856)$ for *perfect mobility*, $(c, d) = (0.285, 0.644)$ for *semi mobility* and $(c, d) = (0.125, 0.334)$ for *weak mobility*. The average number of singletons can as well be described by a linear fit $\bar{N}_0(N) = f + g N$ with $(f, g) = (0.168, 0.739)$ for *perfect mobility*, $(f, g) = (-0.032, 0.441)$ for *semi mobility*, and $(f, g) = (0.495, 0.092)$ for the *weak mobility* setting.

of β on the system on a macro scale. The average size of the largest cluster for different system sizes has the form

$$\bar{k}_{\max}(N) = a N^b \tag{7.8}$$

with $(a, b) = (2.104, 0.121)$ for the *perfect mobility* setting, $(a, b) = (3.49, 0.122)$ for *semi mobility* and $(a, b) = (5.81, 0.132)$ for the *weak mobility* setting. The average number of clusters and the average number of isolated nodes are linear function in N. The exact fits can be found in the caption of Fig. 7.6. Generally, the more immobile the system becomes the fewer clusters and the fewer isolated nodes can be found. Decreasing β does not only slow down the process of agent's withdrawal from existing clusters, it also slows down the addition process. This is evident from Fig. 7.6c. The number of isolated nodes becomes very small compared to the system size, which is due to the lower probability that nodes leave clusters. As a consequence, since only isolated nodes can join existing clusters, this process is also slowed down. Fig. 7.7 illustrates this directly. Fig. 7.8b il-
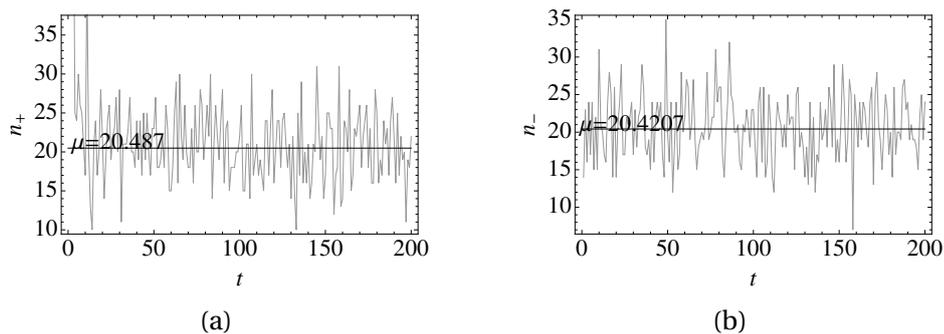

Figure 7.7: Number of additions (a) and removals (b) per time step. Results are obtained in a system with $N = 10,000$ in the strongly biased setting. The average is obtained after $T = 1,000$ time steps.

lustrates the relationship between an agent's fitness and its involvement in joining and withdrawing activities. Agents with low fitness are very often moving out of clusters, but rarely move in. This occurs because the entry barrier for other agents who decide to initiate a new cluster is very low. However, if the cluster does not grow beyond size 2, the low fitness agent has to withdraw in the same time step from the newly built cluster. Generally, joining activity is increasing with fitness, but slows down for very high fitness. This is because high fitness nodes represent seeds of existing clusters and once fixed inside, never leave. However, the level of activity is sufficiently high, because these agents are left alone more frequently, which makes it then necessary for these agents to join new

clusters.

Generally, fitness translates well into ranks. Fig. 7.8c shows the relationship between these two quantities. The plot shows that there is a phase separation. While agents with high fitness can possibly have any rank, since they might be isolated or at the top of the largest cluster, it is impossible for low fitness nodes to stay atop of a large cluster. Herein lies also the reason for the relative small size of the largest cluster. While the present mechanism favours the selection of large clusters, joining them becomes increasingly difficult, since the joining agent must rank somewhere above half of the cluster size. Another view on this phenomenon is given in Fig. 7.9. Fig. 7.9 illustrates the relation-
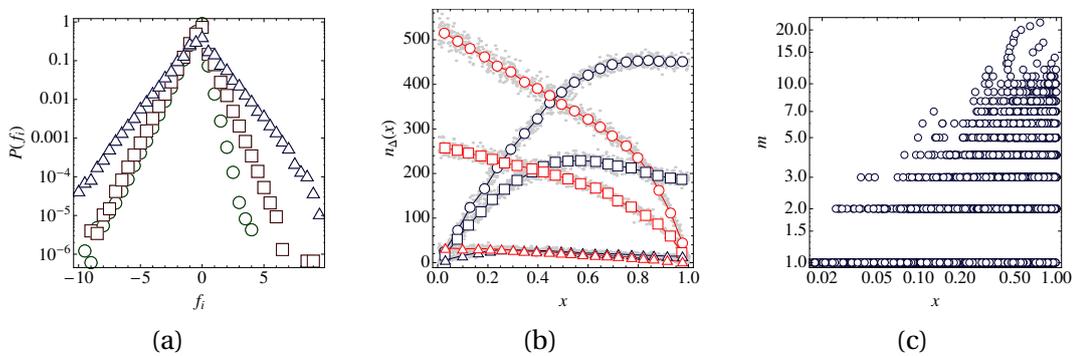

Figure 7.8: Various microscopic indicators in the system: (a) Distribution of the field: $f_{m_i,k} = m_i - (k+1)/2$ at the time of action for the three different scenarios. (b) Agents' activity vs. fitness. The red (decaying) traces are the number of withdrawal steps that an agent with a given fitness was involved in. The blue (increasing) traces are the number of joining events agents with given fitness were involved in. (c) Agents' ranks vs. fitness. Each datapoint is a (rank,fitness)-pair of an agent in the strongly biased system.

ship between fitness and rank when an agent is involved in an event, that is when it either leaves or joins a cluster. The contour plot shows that high fitness nodes are rarely involved in joining activities. This has already been shown in Fig. 7.8b. The highest activity occurs at low ranks. This is due to the low entry barriers for small clusters. The likelihood to rank high in a large cluster is for a low fitness node less than the likelihood to come out highest in a cluster of size 2. Fig. 7.9b shows the relationship between ranking and fitness at the time of withdrawal from existing clusters. It illustrates that highly fit agents, at the top of their cluster, are almost never involved in withdrawals. Most of the activity occurs at the lower end of the ranking scale. For very low rank, this is even indiscriminate for fitness. In the case of high fitness, these are clusters made of two very high fit nodes, where the lower ranked one needs to leave, since these clusters do not

attract other nodes, which could secure the seed and initiate further growth.

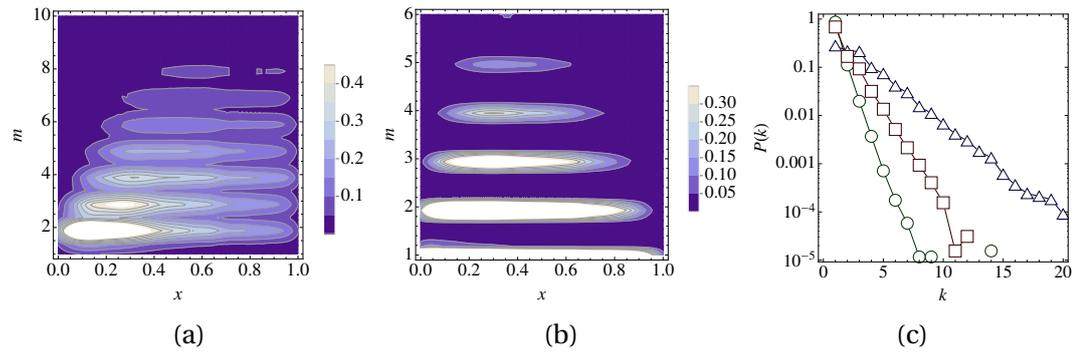(a)   (b)   (c)

Figure 7.9: (a)-(b) Fitness vs. rank when a node becomes active in a system with 10,000 agents after 1,000 time steps for the strongly biased setting. Subfigure (a) shows the relationship on time of joining a cluster, (b) depicts the relationship on withdrawal. (c) Distribution of cluster sizes with $N = 10,000$ in a log-linear plot for the different settings: strong immobility ($\triangle$), weak immobility ($\square$) and no immobility ($\bigcirc$)

Fig. 7.9c illustrates the cluster size distribution. As argued already above, the domain of these distributions is extremely short, compared to the size of the system. A reliable fit of these distributions is not possible because of the very short domain. From the plot, it can be inferred that the distribution is heavier than an exponential.The decay appears over a very short range linear in a log-log scaled plot. These type of mixed exponential, with power-law tail type distributions have been found elsewhere, for instance in wealth distributions [44]. However, these results cannot be reliably inferred here from the numerical results.

4   Concluding Remarks

A cluster formation system, based on individual preferences has been analysed in this chapter. Curiously, the system failed to reach a steady state cluster size distribution. The size of the largest cluster fluctuates around its mean. These fluctuations become stronger the more the system is put into a regime of immobility. It has been illustrated here that this is a direct consequence of the closeness of the system and its compact domain of fitness. The system needs to emit a certain number of high fitness agents into isolation, so that these can seed new growing clusters.

The cluster size distribution decays with a power-law tail. However, because of the finite
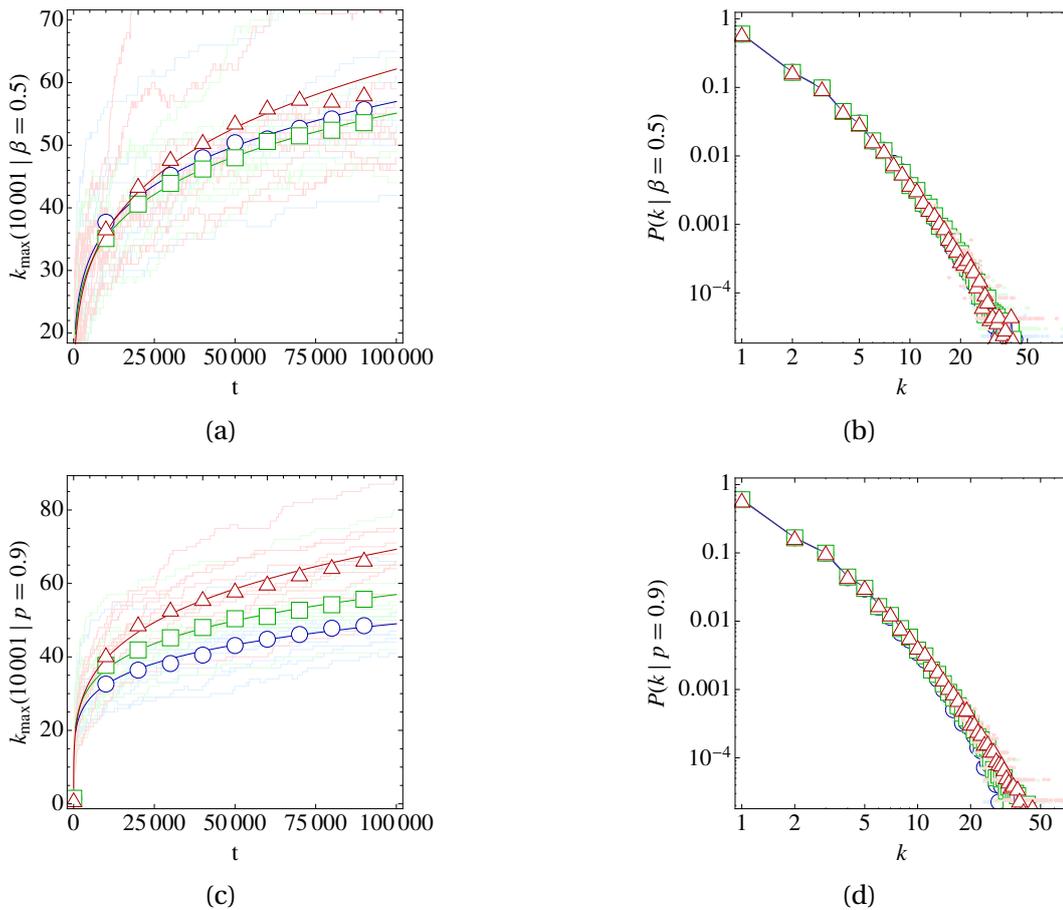
(a)

(b)

(c)

(d)

Figure 7.10: Results for the system with growth after 100,000 time steps. Development of the largest cluster and the cluster size distribution with $\beta = 0.5$ and $q = 0.9$ ($\bigcirc$), $q = 0.7$ ($\square$), and $q = 1/2$ ($\triangle$) in panels (a),(b). For $q = 0.9$ and $\beta = 1$ ($\bigcirc$), $\beta = 1/2$ ($\square$), and $\beta = 1/20$ ($\triangle$) in panels (c),(d).

system size, the support is very compressed and therefore the power-law exponent extremely high. Due to the lack of broadness of the distribution, reliable estimates of these exponents cannot be given.

It has been shown that the system exhibits larger cycles and a broader distribution when its updating process is slowed down. The ultimately slowed down model, in which just one agent can move per time step, that was discussed in section 3.1, shows similar behaviour as the full model.

Another possible variation is the introduction of growth. Adding single nodes to the system makes only sense in a comparison to the slowed down alternative, because of the different time scale. For the growing model, a probability q is introduced. At every time step, with probability q a new agent is joined to the system. The rules of joining are the

same as in the static case. The new agent attempts to join a cluster preferentially with size. If it ranks to low inside the chosen cluster, the new agent initiates a new cluster. If it ranks sufficiently high, it will join the cluster. With probability $1 - q$ an internal update step occurs. The rules are the same here as in the slowed down static version, that is discussed before. Numerical results for this variation can be found in Fig. 7.10.

The two variations of the original model lead to two conclusions. On one side, it is clear that the result obtained in the body of this chapter is robust against a change of the time scale in the system. On the other side, the dynamic variation confirms the consequences from findings in the main model. It has been found here that slowing down the withdrawal of agents from clusters is not sufficient to build large clusters, because of the fixed size of the system. When agents stop withdrawing, there are not enough agents to move into existing clusters to increase their size. Adding an external source of agents can solve this issue. However, the system size still remains relatively small compared to real-world systems.

All these results show that cluster size growth cannot be stimulated by solely skewing the probabilities of leaving and joining clusters within the present framework. Growth is a necessary ingredient to facilitate the composition of large clusters because the system needs a steady influx of new agents. This is because high fitness agents often occupy clusters of small size which are then rarely joined because of the high entry barrier. But growth alone does not lead to large clusters either. The decomposition of existing clusters occurs at a very long time scale in the strongly biased system, but this decomposition is still too fast compared to the process of agent additions to existing clusters. Therefore further model ingredients are necessary to facilitate the coagulation of hub-like clusters that occupy large fractions of the system.

In summary, this chapter has opened an investigation into the formation of clusters based on individual preferences. It has been pointed out here that the investigation into cluster formation without underlying geometry is very different from the investigation into clusters in the context of networks. However, this area of research is an interesting branch, as there are many structures that are not set in a natural geometry, but rather just

form collections of individuals. The motivating example for the present investigation are firms, which can be represented by collections of agents. But also other group formations, such as the composition of discussion groups on the internet can be understood in this way. In that case people do not establish connections to every other person in the network, but rather add themselves to the collective of a whole group. Another means of modelling these aspects are bipartite networks, where one partition represents clusters and the other partition represents agents. However, while the agent's partition follows simple rules, the cluster side clearly exhibits very complex behaviour, as nodes emerge and vanish based on the rules. A deep investigation into the consequences of various attachment rules is left for future research.

Computational Methodology

Many of the results in this thesis are backed by numerical simulations of the systems in question. The main difficulty of these simulations is to formulate not only the underlying topology but also its function in the most efficient way. Efficient formulation is essential. This is best illustrated with an example. Consider for instance a network with $N = 10^5$ many nodes. The most intuitive description of a binary network is an adjacency matrix, filled with ones and zeros. These entries can be represented as boolean variables. Using Java, the adjacency matrix will need at least 10 Gigabytes of memory, if one assumes that a boolean can be stored in one byte[1]. To overcome this immense need of storage, there are a number of possible solutions. One way is to represent the network as an edge-list, rather than as an adjacency matrix. If the network is sparse, then this list will be relatively short, much shorter than N^2. Alternatively, there exist a number of efficient approaches in the numerical methods literature. One approach that has been considered for the present thesis is the Compressed Row Storage approach, where instead of the entire adjacency matrix, only the non-zero elements and their positions are stored.

The storage problem becomes even more cumbersome if edges need to be sampled according to some criterion that depend on the fitness of adjacent nodes. Consider again the network with $N = 10^5$ nodes. Placing all possible edges together with a probability

---

[1] Notice that, depending on the Java virtual machine, the overhead for booleans can vary. Nevertheless 1 byte as a lower bound is a good estimator. Other languages might store the one bit of information that a boolean carries more efficiently.

depending on the adjacent fitness into a sampler uses approximately $(10^5)^2/2 \cdot 64\text{bit} = 40\text{GB}$ of memory in an undirected network using the full adjacency matrix. This is prohibitive much memory for most work stations.

Another issue is the speed of computations. Consider for example the time varying network. In each network instance, the entire network need to be sampled. Each node can activate and connect to various positions inside the network. After each instance, the adjacency matrix is reset and a new instance computed. This process is repeated many times and finally all of the collected results are averaged over many network realisations. In order to compute these aspects efficiently, language and internal data structures need to be chosen carefully. Consider for instance the summing of $M = 10^7$ random numbers. Compiled languages differ substantially from script languages. But even for compiled languages exists a huge difference of computation times depending on the exact implementation. Using Java on a Mac with an Intel i7 processor with 2.8Ghz, two multithreading cores, and 16GB of memory, the summation takes 15 milliseconds when an `array` is used. The same computation takes 26 milliseconds when the values are stored in an `ArrayList`, rather than an `array`. `array` and `ArrayList` are just two of many different data structures that can be chosen to store the 10^7 random numbers, but the example illustrates clearly the importance of a well chosen implementation of the problem at hand. For script languages, like Mathematica for example, the problem becomes even more severe. A naive implementation of the problem uses the inbuilt `Sum[]`-function, for which timing scales worse than with a power in the length of the vector and uses already more than one second for 10^4 items. Fig. A.1 illustrates the relationship between the size of the vector and the computing time. A more native implementation, utilising the core of the language, namely lists, uses `Apply[]` to compute the sum. `Apply[plus,...]` needs 2.11sec. to solve the problem. That is more than 1000-times slower than the Java implementation on `array`'s.

Depending on the task at hand either one or all of Java, R, and Mathematica have been used to compute the results in this thesis. While Java as a compiled language is extremely fast, compared to the latter two, R and Mathematica are very versatile languages for prototyping. Both have inbuilt network data structures and the main properties such as the
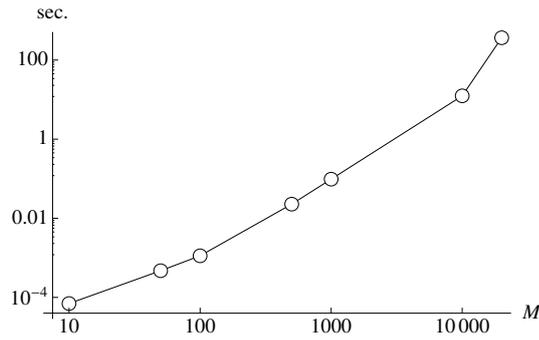
Figure A.1: Relationship of computing times of the sum of real numbers and the number of random numbers M in Mathematica using a naive approach.

degree distribution are already implemented. Furthermore, R is a very rich language to conduct data analysis and has been used to calculate results from the simulations that were run on Java. Mathematica has mainly been used to do computer algebra, such as finding solutions of non-linear equations and numerical computations of integrals. Behind that, most of the figures in this thesis are produced with Mathematica.

All random numbers were generated by inverse transform sampling, using the inbuilt algorithms to produce standard uniform random numbers.

Bibliography

[1] Abe, S. and Thurner, S. Complex networks emerging from random graphs: analytic formula for the hidden variable distribution. *Physical Review E*, 72:036102, 2005.

[2] Albert, R., Albert, I., and Nakarado, G. L. Structural vulnerability of the north american power grid. *Phys. Rev. E*, 69:025103, 2004.

[3] Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.

[4] Albert, R., Jeong, H., and Barabasi, A.-L. Diameter of the world-wide web. *Nature (London)*, 401(130), 1999.

[5] Albert, R., Jeong, H., and Barabási, A.-L. Error and attack tolerance of complex networks. *Nature*, 406:378–381, 2000.

[6] Alchian, A. A. and Demsetz, H. Production, information costs, and economic organization. *The American Economic Review*, 62(5):777–795, 1972.

[7] Appel, K. and Haken, W. The solution of the four-color-map problem. *Scientific American*, 237:108–121, 1977.

[8] Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[9] Barabâsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3):590–614, 2002.

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

[11] Barigozzi, M., Fagiolo, G., and Mangioni, G. Identifying the community structure of the international-trade multi-network. *Physica A: Statistical Mechanics and its Applications*, 390(11):2051–2066, 2011.

[12] Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.

[13] Barrat, A., Barthlemy, M., and Vespignani, A. *Dynamical processes on complex networks.* Cambridge University Press, 2008.

[14] Barrat, A. and Weigt, M. On the properties of small-world network models. *The European Physical Journal B - Condensed Matter and Complex Systems*, 13(3):547–560, 2000.

[15] Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. Debtrank: too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2(541), 2012.

[16] Bedogne, C. and Rodgers, G. J. Complex growing networks with intrinsic vertex fitness. *Physical Review E*, 74(4):046115, 2006.

[17] Bernoulli, D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad. Roy. Sci.(Paris) avec Mém. des Math. et Phys. and Mém*, pages 1–45, 1760.

[18] Bhattacharya, K., Mukherjee, G., Saramäki, J., Kaski, K., and Manna, S. S. The international trade network: weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(02):P02002, 2008.

[19] Bianconi, G. and Barabási, A.-L. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436, 2001.

[20] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175–308, 2006.

[21] Boguñá, M. and Pastor-Satorras, R. Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68:036112, 2003.

[22] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

[23] Bullmore, E. and Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198, 2009.

[24] Caldarelli, G., Capocci, A., De Los Rios, P., and Muñoz, M. Scale-free networks from varying vertex intrinsic fitness. *Physical review letters*, 89(25):258702, 2002.

[25] Callaway, D. S., Newman, M. E. J., Strogatz, S. H., and Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85:5468–5471, 2000.

[26] Castellano, C., Vilone, D., and Vespignani, A. Incomplete ordering of the voter model on small-world networks. *EPL (Europhysics Letters)*, 63(1):153, 2003.

[27] Cauchemez, S., Bhattarai, A., Marchbanks, T. L., Fagan, R. P., Ostroff, S., Ferguson, N. M., Swerdlow, D., and the Pennsylvania H1N1 working group. Role of social networks in shaping disease transmission during a community outbreak of 2009 h1n1 pandemic influenza. *Proceedings of the National Academy of Sciences*, 108(7):2825–2830, 2011.

[28] Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., and Faloutsos, C. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4):1, 2008.

[29] Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J. Can cascades be predicted? In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 925–936. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2014.

[30] Chiarucci, R., Ruzzenenti, F., and Loffredo, M. I. Detecting spatial homogeneity in the world trade web with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 401(0):1 – 7, 2014.

[31] Clauset, A., Shalizi, C., and Newman, M. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[32] Clifford, P. and Sudbury, A. A model for spatial conflict. *Biometrika*, 60(3):pp. 581–588, 1973.

[33] Coase, R. H. The nature of the firm. *economica*, 4(16):386–405, 1937.

[34] Cohen, R., Erez, K., ben Avraham, D., and Havlin, S. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628, 2000.

[35] Colizza, V., Barrat, A., Barthélemy, M., and Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, 2006.

[36] Crucitti, P., Latora, V., and Marchiori, M. A topological analysis of the italian electric power grid. *Physica A: Statistical Mechanics and its Applications*, 338(1–2):92 – 97, 2004. Proceedings of the conference A Nonlinear World: the Real World, 2nd International Conference on Frontier Science.

[37] De Masi, G., Iori, G., and Caldarelli, G. Fitness model for the italian interbank money market. *Phys. Rev. E*, 74:066112, 2006.

[38] D'Hulst, R. and Rodgers, G. J. Exact solution of a model for crowding and information transmission in financial markets. *International Journal of Theoretical and Applied Finance*, 3(04):609–616, 2000.

[39] D'hulst, R. and Rodgers, G. J. Business size distributions. *Physica A: Statistical Mechanics and its Applications*, 299(1):328–333, 2001.

[40] D'Hulst, R. and Rodgers, G. J. Models for the size distribution of businesses in

a price driven market. *The European Physical Journal B-Condensed Matter and Complex Systems*, 21(3):447–453, 2001.

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

[42] Dorogovtsev, S. N. and Mendes, J. F. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.

[43] Drăgulescu, A. and Yakovenko, V. Exponential and power-law probability distributions of wealth and income in the united kingdom and the united states. *Physica A: Statistical Mechanics and its Applications*, 299(1):213–221, 2001.

[44] Dragulescu, A. A. Applications of physics to economics and finance: money, income, wealth, and the stock market. *arXiv preprint cond-mat/0307341*, 2003.

[45] Eguíluz, V. M. and Zimmermann, M. G. Transmission of information and herd behavior: An application to financial markets. *Phys. Rev. Lett.*, 85:5659–5662, 2000.

[46] Erdős, P. and Rényi, A. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 5:17–61, 1960.

[47] Ergün, G. and Rodgers, G. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1):261–272, 2002.

[48] Euler, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.

[49] Fagiolo, G. Directed or undirected? a new index to check for directionality of relations in socio-economic networks. *Economics Bulletin*, 3(34):1–12, 2006.

[50] Fagiolo, G., Reyes, J., and Schiavo, S. World-trade web: Topological properties, dynamics, and evolution. *Phys. Rev. E*, 79:036115, 2009.

[51] Fagiolo, G., Reyes, J., and Schiavo, S. The evolution of the world trade web: a weighted-network analysis. *Journal of Evolutionary Economics*, 20(4):479–514, 2010.

[52] Faloutsos, M., Faloutsos, P., and Faloutsos, C. On power-law relationships of the internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4):251–262, 1999.

[53] Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., and Mo, H. World trade flows: 1962-2000. Technical report, National Bureau of Economic Research, 2005.

[54] Fronczak, A. and Fronczak, P. Statistical mechanics of the international trade network. *Phys. Rev. E*, 85:056113, 2012.

[55] Garas, A., Argyrakis, P., Rozenblat, C., Tomassini, M., and Havlin, S. Worldwide spreading of economic crisis. *New Journal of Physics*, 12(11):113043, 2010.

[56] Garlaschelli, D., Di Matteo, T., Aste, T., Caldarelli, G., and Loffredo, M. I. Interplay between topology and dynamics in the world trade web. *The European Physical Journal B*, 57(2):159–164, 2007.

[57] Garlaschelli, D. and Loffredo, M. I. Fitness-dependent topological properties of the world trade web. *Phys. Rev. Lett.*, 93:188701, 2004.

[58] Garlaschelli, D. and Loffredo, M. I. Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications*, 355(1):138–144, 2005.

[59] Garlaschelli, D. and Loffredo, M. I. Effects of network topology on wealth distributions. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224018, 2008.

[60] Gleditsch, K. S. Expanded trade and gdp data. *Journal of Conflict Resolution*, 46(5):712–724, 2002.

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

[62] Guare, J. *Six degrees of separation: A play*. Random House LLC, 1990.

[63] Guimerà, R., Mossa, S., Turtschi, A., and Amaral, L. N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 2005.

[64] Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J., Cusick, M. E., Roth, F. P., et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, 430(6995):88–93, 2004.

[65] Helbing, D., Brockmann, D., Chadefaux, T., Donnay, K., Blanke, U., Woolley-Meza, O., Moussaid, M., Johansson, A., Krause, J., Schutte, S., et al. Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics*, pages 1–47, 2014.

[66] Hethcote, H. W. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

[67] Hill, S. A. and Braha, D. Dynamic model of time-dependent complex networks. *Phys. Rev. E*, 82:046105, 2010.

[68] Holme, P. and Saramäki, J. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.

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

[70] Huberman, B. A. and Adamic, L. A. Internet: Growth dynamics of the world-wide web. *Nature*, 401(6749):131–131, 1999.

[71] Iribarren, J. L. and Moro, E. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.*, 103:038702, 2009.

[72] Jeong, H., Néda, Z., and Barabási, A. Measuring preferential attachment in evolving networks. *EPL (Europhysics Letters)*, 61(4):567, 2003.

[73] Kali, R. and Reyes, J. The architecture of globalization: a network approach to international economic integration. *Journal of International Business Studies*, 38(4):595–620, 2007.

[74] Kali, R. and Reyes, J. Financial contagion on the international trade network. *Economic Inquiry*, 48(4):1072–1101, 2010.

[75] Knuth, D. *The Art of Computer Programming*, volume 1. MA: Addison-Wesley, 1968.

[76] Krebs, V. Uncloaking terrorist networks. *First Monday*, 7(4), 2002.

[77] Kretschmar, M. and Morris, M. Measures of concurrency in networks and the spread of infectious disease. *Mathematical biosciences*, 133(2):165–195, 1996.

[78] Kuperman, M. and Abramson, G. Small world effect in an epidemiological model. *Phys. Rev. Lett.*, 86:2909–2912, 2001.

[79] Lawless, M. Deconstructing gravity: trade costs and extensive and intensive margins. *Canadian Journal of Economics/Revue canadienne d'économique*, 43(4):1149–1172, 2010.

[80] Lèbre, S., Becq, J., Devaux, F., Stumpf, M., and Lelandais, G. Statistical inference of the time-varying structure of gene-regulation networks. *BMC systems biology*, 4(1):130, 2010.

[81] Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Åberg, Y. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[82] Liu, S.-Y., Baronchelli, A., and Perra, N. Contagion dynamics in time-varying metapopulation networks. *Phys. Rev. E*, 87:032805, 2013.

[83] Meloni, S., Perra, N., Arenas, A., Gomez, S., Moreno, Y., and Vespignani, A. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.*, 1, 2011.

[84] Mitzenmacher, M. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.

[85] Molloy, M. and Reed, B. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[86] Moore, C. and Newman, M. E. J. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61:5678–5682, 2000.

[87] Moore, C. and Newman, M. E. J. Exact solution of site and bond percolation on small-world networks. *Phys. Rev. E*, 62:7059–7064, 2000.

[88] Moretti, P., Liu, S., Baronchelli, A., and Pastor-Satorras, R. Heterogenous mean-field analysis of a generalized voter-like model on networks. *The European Physical Journal B*, 85(3):1–6, 2012.

[89] Newman, M. E. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[90] Newman, M. E. J. Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E*, 64:016131, 2001.

[91] Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.

[92] Newman, M. E. J. Component sizes in networks with arbitrary degree distributions. *Phys. Rev. E*, 76:045101, 2007.

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

[94] Organisation for Economic Cooperation and Development. International trade by commodities statistics. Technical report, Mimas, University of Manchester, 2010. DOI: http://dx.doi.org/10.5257/oecd/itcs/2007.

[95] Padgett, J. F. and Ansell, C. K. Robust action and the rise of the medici, 1400-1434. *American journal of sociology*, pages 1259–1319, 1993.

[96] Pan, R. K. and Saramäki, J. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E*, 84:016105, 2011.

[97] Pastor-Satorras, R. and Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.

[98] Peng, X.-L., Xu, X.-J., Fu, X., and Zhou, T. Vaccination intervention on epidemic dynamics in networks. *Phys. Rev. E*, 87:022813, 2013.

[99] Penzar, D. About modelling of complex networks with applications to terrorist group modelling. *Interdisciplinary Description of Complex Systems*, 3(1):27–43, 2005.

[100] Perra, N., Gonçalves, B., Pastor-Satorras, R., and Vespignani, A. Activity driven modeling of time varying networks. *Scientific Reports*, 2(469), 2012.

[101] Picciolo, F., Squartini, T., Ruzzenenti, F., Basosi, R., and Garlaschelli, D. The role of distances in the world trade web. In *The 8th International Conference on Signal Image Technology and Internet Based Systems (SITIS 2012)*, pages 784–792. IEEE, Conference Publishing Services, Sorrento, Italy, 2012.

[102] Price, D. d. S. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.

[103] Price, D. d. S. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27:292–306, 1976.

[104] Rodgers, G. and Zheng, D. A herding model with preferential attachment and fragmentation. *Physica A: Statistical Mechanics and its Applications*, 308(1):375–380, 2002.

[105] Roukny, T., Bersini, H., Pirotte, H., Caldarelli, G., and Battiston, S. Default cascades in complex networks: Topology and systemic risk. *Scientific Reports*, 3, 2013.

[106] Russell, S. and Norving, P. *Artificial Intelligence: A Modern Approach*. Prentice Hall Artificial Intelligence Series. Pearson Education Inc., 2nd edidtion edition, 2003.

[107] Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., and White, D. R. Economic networks: The new challenges. *Science*, 325(5939):422–425, 2009.

[108] Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., and White, D. R. Economic networks: What do we know and what do we need to know? *Advances in Complex Systems*, 12(04n05):407–422, 2009.

[109] Serrano, M. A. and Boguñá, M. Topology of the world trade web. *Phys. Rev. E*, 68:015101, 2003.

[110] Serrano, M. A. and Boguñá, M. Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.*, 97:088701, 2006.

[111] Servedio, V. D., Caldarelli, G., and Butta, P. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Physical Review E*, 70(5):056126, 2004.

[112] Sood, V., Antal, T., and Redner, S. Voter models on heterogeneous networks. *Phys. Rev. E*, 77:041121, 2008.

[113] Sood, V. and Redner, S. Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, 94:178701, 2005.

[114] Spulber, D. F. *The theory of the firm: Microeconomics with endogenous entrepreneurs, firms, markets, and organizations.* Cambridge University Press, 2009.

[115] Squartini, T., Fagiolo, G., and Garlaschelli, D. Randomizing world trade. i. a binary network analysis. *Phys. Rev. E*, 84:046117, 2011.

[116] Squartini, T., Fagiolo, G., and Garlaschelli, D. Randomizing world trade. ii. a weighted network analysis. *Phys. Rev. E*, 84:046118, 2011.

[117] Stehlé, J., Barrat, A., and Bianconi, G. Dynamical and bursty interactions in social networks. *Phys. Rev. E*, 81:035101, 2010.

[118] Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, 122(6):957 – 968, 2005.

[119] Tadić, B. and Rodgers, G. Modelling conflicts with cluster dynamics in networks. *Physica A: Statistical Mechanics and its Applications*, 389(23):5495–5502, 2010.

[120] Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J. J., Balcan, D., Gonçalves, B., Perra, N., Colizza, V., and Vespignani, A. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC medicine*, 10(1):165, 2012.

[121] Vazquez, A., Rácz, B., Lukács, A., and Barabási, A.-L. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.*, 98:158702, 2007.

[122] Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2011.

[123] Wang, Y., Chakrabarti, D., Wang, C., and Faloutsos, C. Epidemic spreading in real networks: an eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34. 2003.

[124] Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[125] Watts, D. J. and Strogatz, S. H. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[126] Weng, L., Flammini, A., Vespignani, A., and Menczer, F. Competition among memes in a world with limited attention. *Sci. Rep.*, 2, 2012.

[127] Wilf, H. S. *generatingfunctionology*. Academic Press, Inc., 1994.

[128] Williamson, O. E. Transaction-cost economics: The governance of contractural relations. *Journal of Law and Economics*, 22(2):233–61, 1979.

[129] Xu, X.-J., Peng, X.-L., Zhang, L.-J., and Yang, G.-H. Effect of fitness on mutual selection in network evolution. *Physics Procedia*, 3(5):1795–1799, 2010.

[130] Xu, X.-J. and Zhang, L.-M. Mutual selection in network evolution: The role of intrinsic fitness. *International Journal of Modern Physics C*, 21(129), 2010.

[131] Zhao, K., Stehlé, J., Bianconi, G., and Barrat, A. Social network dynamics of face-to-face interactions. *Phys. Rev. E*, 83:056109, 2011.

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