INFORMATION HORIZONS IN A COMPLEX WORLD

Martin Rosvall
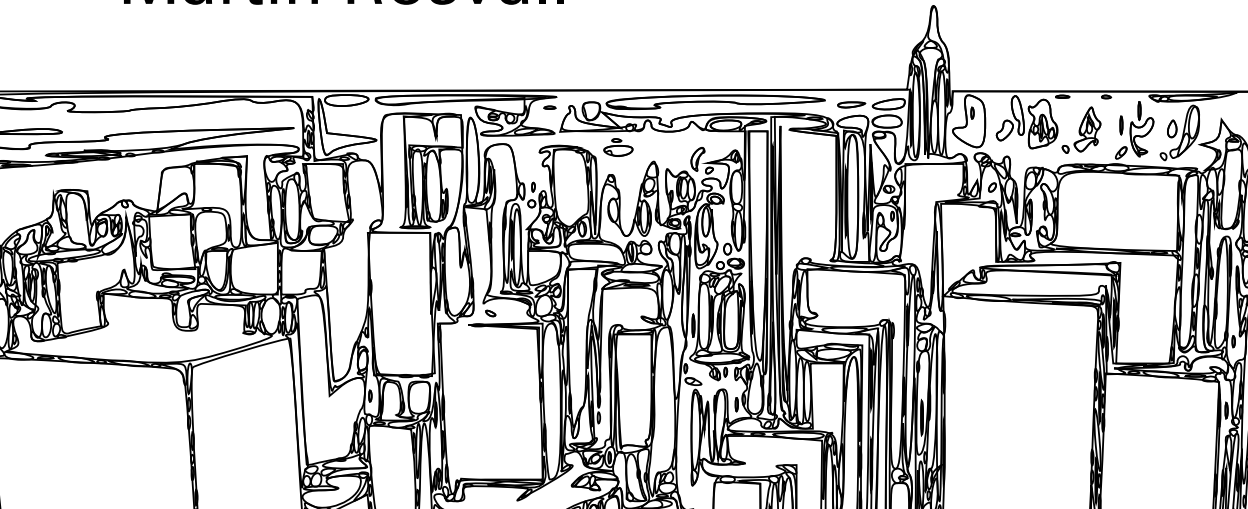
INFORMATION HORIZONS IN A COMPLEX WORLD

Martin Rosvall

Abstract

The whole in a complex system is the sum of its parts, plus the inter-actions between the parts. Understanding social, biological, and economic systems therefore often depends on understanding their patterns of interactions—their networks. In this thesis, the approach is to understand complex systems by making simple network models with nodes and links. It is first of all an attempt to investigate how the communication over the network affects the network structure and, vice versa, how the network structure affects the conditions for communication.

To explore the local mechanism behind network organization, we used simplified social systems and modeled the response to communication. Low communication levels resulted in random networks, whereas higher communication levels led to structured networks with most nodes having very few links and a few nodes having very many links. We also explored various models where nodes merge into bigger units, to reduce communication costs, and showed that these merging models give rise to the same kind of structured networks.

In addition to this modeling of communication networks, we developed new ways to measure and characterize real-world networks. For example, we found that they in general favor communication on short distance, two-three steps away in the network, within what we call the information horizon.

Sammanfattning

Helheten i ett komplext system är mer än summan av dess delar, då den även inbegriper interaktionerna mellan dem. Att studera sociala, biologiska och ekonomiska system blir därför ofta en fråga om att förstå deras interaktionsmönster, d.v.s. deras nätverk av noder och länkar. Med utgångspunkt i enkla nätverksmodeller undersöker avhandlingen i huvudsak hur kommunikation i nätverk påverkar nätverksstrukturen och, vice versa, hur nätverksstrukturen påverkar villkoren för kommunikation.

Vi utforskade mekanismerna bakom hur nätverk är organiserade genom att modellera effekten av kommunikation i förenklade sociala system. En låg kommunikationsnivå visade sig ge upphov till kaotiska nätverk där ingen nod i princip hade fler länkar än någon annan. En hög kommunikationsnivå resulterade däremot i strukturerade nätverk, med några få centrala noder med många länkar, medan flertalet noder var perifera med enbart några få länkar. Det visade sig också att alla aktörer i nätverket gynnades av kommunikation, även när den var ojämnt fördelad. Kvaliteten på kommunikationen, d.v.s. informationens giltighet, var också avgörande för vilka positioner som gynnades i ett nätverk, vilket vi visade genom att studera aktörer som spred falsk information.

Eftersom effektiv kommunikation är en viktig del i många nätverk betraktar vi utvecklingen av dem som en optimeringsprocess. Varje kommunikationshandling mellan noderna tar tid och genom att slå sig samman till större enheter begränsas dessa kostnader och gör nätverket effektivare. Dessa s.k. sammanslagningsmodeller gav upphov till samma typ av strukturerade nätverk som ovan.

Genom att utveckla olika sätt att mäta nätverksstrukturer visade vi bland annat att många verkliga system främjar kommunikation över korta avstånd, två-tre steg bort i nätverket, innanför det vi kallar informationshorisonten. Vi uppskattade också den mängd information som krävs för att orientera sig i städer, och fann att det är lättare att hitta i moderna, planerade städer än i äldre städer som utvecklats under lång tid.

Papers

This thesis is based on the following papers (reprinted with kind permission by the publishers):

I   Axelsen, J., Bernhardsson, S., Rosvall, M., Sneppen, K., and Trusina, A., *Degree landscapes in scale-free networks*, Phys. Rev. E (in Press).

II   Minnhagen, P., Rosvall, M., Sneppen, K., and Trusina, A. (2004), *Self-organization of structures and networks from merging and small-scale fluctuations*, Physica A: Statistical Mechanics and its Applications **340**, 725–732.

III   Sneppen, K., Rosvall, M., Trusina, A., and Minnhagen, P. (2004), *A simple model for self organization of bipartite networks*, Europhys. Lett. **67**, 349–354.

IV   Sneppen, K., Trusina, A., and Rosvall, M. (2005), *Hide and seek on complex networks*, Europhys. Lett. **69**, 853–859.

V   Rosvall, M., Trusina, A., Minnhagen, P., and Sneppen, K. (2005), *Networks and cities: An information perspective*, Phys. Rev. Lett. **94**, art. no. 028701.

VI   Trusina, A., Rosvall, M., and Sneppen, K. (2005), *Communication boundaries in networks*, Phys. Rev. Lett. **94**, art. no. 238701.

VII   Rosvall, M., Minnhagen, P., and Sneppen, K. (2005), *Navigating networks with limited information*, Phys. Rev. E **71**, art. no. 066111.

VIII   Rosvall, M., Grönlund, A., Minnhagen, P., and Sneppen, K. (2005), *Searchability of networks*, Phys. Rev. E **72**, art. no. 046117.

IX Rosvall, M. and Sneppen, K. (2003), *Modeling dynamics of information networks*, Phys. Rev. Lett. **91**, art. no. 178701.

 X Rosvall, M. and Sneppen, K. (2006), *Self-assembly of information in networks*, Europhys. Lett. **74**, 1109–1115.

XI Rosvall, M. and Sneppen, K. (2006), *Modeling self-organization of communication and topology in social networks*, Phys. Rev. E **74**, art. no. 016108.

Other papers by the author:

• Sneppen, K., Trusina, A., and Rosvall, M. (2005), *Measuring information networks*, Pramana-J. Phys. **6**, 1121–1125.

• Rosvall, M. and Sneppen, K., *Networks and our limited information horizon*, Int. J. Bifurcat. Chaos (in Press).

Preface

"The book that nobody read" was written more than 450 years ago (Gingerich, 2004). Despite the title it was not written by a Ph.D. student. Instead, the book was Nicolaus Copernicus' work of life *De revolutionibus orbium coelestium*[1]. Copernicus moved the center of the universe from the earth to the sun in this epoch-making work. The result was not a system that gave better predictions of the real world than the old Ptolemaic viewpoint, but a victory of the simplicity. He replaced a monster of independently moving units in the geocentric universe of Ptolemy by a system of interconnected planets all circulating around the sun. To conciliate the Catholic Church he had to write that he was only hypothesizing about the universe and that his system did not correspond to reality. We now know that it really does, but whether Copernicus actually believed in the new simple conception of the world that he created is not clear. Nevertheless, it was the strive for simplicity that took him there.

My ambition to try to understand the world through simple models does not originate in a fascination for the heavenly spheres. Instead, I prefer to look at earthly things, like ants building their nests. Ants, and not an ant, since neither only one ant on the ground, nor one sphere in the sky makes a system. Nine planets, roughly one hundred moons, and a sun make a solar system; hundreds of workers and alates[2] and a queen make a colony. However, here end the similarities between the two systems. When the solar system is not much more than some objects circulating around each other (from an astronomer's point of view), the colony is apparently much more than a couple of ants. It is a superorganism that builds its nest without a leader who commands the workers what to do. An ant is for me a player with a limited set of instructions, and the colony a game with several interacting players—to some extent like our human society. In other

---

[1] *On the revolutions of the heavenly spheres*, often referred to as *De revolutionibus*, was published the same year as Nicholas Copernicus (1473-1543) died.

[2] Alates are male ants with wings.

words, this is not unique for ants, but permeates many systems in nature. There-
fore, I think that an ant colony serves as a good picture of what a *complex system*
is, emphasizing that a complex system is simple *simple* rather than *complicated*.
It just happens to be played on the globe orbiting the sun.

The aim of this thesis is not to once again move the center of the universe,
but to focus on the difference between Copernicus *system* and the *independent
units* of Ptolemy. The story will be about the difference between the whole and
the sum of its parts—the interactions.

In the end, "the book that nobody read" was in fact read by Kepler, Brahe, da
Vinci and others, who were all affected by the content of *De revolutionibus*. Even
though your name is unknown to me, I hope that I can move the information
horizon of your complex world—if only slightly.

Acknowledgments

This thesis is about networks and is also a product of one. First, I thank my supervisors Petter Minnhagen and Kim Sneppen for their continuous support. Petter invited me to the network of science and has always encouraged me to keep my link to skiing, which has meant everything to me. My best education is so far the plenty of afternoons spent in Kim's office together with a Classic Cola and unrestrained discussions about just anything. They, together with my inspiring co-authors on the papers: Jacob Bock Axelsen, Sebastian Bernhardsson, Andreas Grönlund, and Ala Trusina, form the center of my scientific network.

On the journey toward the thesis, I have also got invaluable help from Ann-Charlott Dalberg and her Danish counterpart Helle Kiilerich. They form the hubs of my Swedish and Danish base camps. Therefore, I let them represent all colleagues who have supported me at the Department of Physics in Umeå and at NORDITA/The Niels Bohr Institute in Copenhagen.

A number of people generously spent time on reading the manuscript and their critical comments made this thesis much better. Therefore, I offer sincere thanks to: Lars Fransson, Karin Jonsell, Petter Holme, Gunnel and Ola Rosvall, and Marit Sigurdson.

A special thanks goes to Bernt Johansson who kindly provided me with electricity during the summer 2006, when most of this thesis was written in our cottage in the southern part of Gotland. I also thank Joel Klehn and Ylva Rosvall who gave me an extensive lecture on trends in jeans fashion.

Finally the warmest thanks to Marit Sigurdson again, for always being on my side.

Contents

Chapter 1

Introduction

W hat wrote "To be, or not to be: that is the question?" The brain that put words in Hamlet's mouth consisted of roughly 100 billion neurons, the same number as in your or my brain. But when did we write our last sonnets? This thesis is about how the whole sometimes is greater (and sometimes is less) than the sum of its parts.[1] The essence of network science is that the whole is the sum of its parts plus the interactions between them.

1.1 Why networks?

Complex networks are neither magnetic or charged, nor do they emit light. They are not made of any exotic matter, they are not radioactive and you will never hear: Complex networks offer the world a secure energy supply! So why am I, as a physicist, interested in modeling complex networks?

Networks make it possible to characterize the complex systems of our world, in the same way as a map describes the surrounding landscape. A network is a map of interactions, for example of communication in a social system. As *Homo loquens*, the speaking man (Fry, 1977), communication is fundamental in our society. If we want to understand the complexity in different systems in the world, we must first understand their basic interaction patterns. These networks are often neither regular lattices[2], nor are all units connected randomly—the interaction patterns are complex.

---

[1]It is perhaps more correct to say that the whole rarely is greater, but often is less than the sum of its parts. However, my point is that the whole is not equal to the sum of its parts.

[2]A periodic structure like a checkboard (see Figure 3.1(a)).

Complex, because a *complex system* is made up of a large number of compo-
nents, or *agents*, interacting in such a way that their collective behavior is not
a simple combination of their individual behavior (Newman, 2002b). Craig
Reynolds has expressed it strikingly as "A flock is not a big bird" (Waldrop,
1992). Moreover, the interactions in a complex system often go beyond the sys-
tem itself to make it adaptive to an ever changing environment. On the other
hand, a *simple system* is a system with a limited set of interacting units with a be-
havior that can be described by laws, for example a pendulum or a bouncing ball
(Amaral and Ottino, 2004). Further, a *complicated system* is a system with a large
set of components, each with exact roles that also together act under absolute
laws, for example billiards or cars.

When I started by asking *what* instead of *who* wrote Hamlet's famous ques-
tion, I did not intend to confuse you about who actually wrote Shakespeare's
plays. Instead I wanted to emphasize that he was not isolated. Shakespeare's net-
work went beyond the 100 billion neurons connected by an even larger number
of physical connections inside his brain. Thereby connecting virtually to cul-
tural, historical and social entities like myths, books, and individuals—that in
their turn had their connections. Hamlet would probably have used a different
vocabulary if Shakespeare's *network* had been different.

Cultural, sociological, economic, political, biological, and ecological systems
all evolve under the influence of numerous components. The challenge is: how
simple can we make the rules of the individuals, constrained by interactions on
a network, to reproduce the observed collective behavior in the real world? In
particular, how should this network of interactions be constructed?

I now again turn to the analogy with maps and use a passage from *Sylvie and
Bruno Concluded* (Carroll, 1893).

> "What a useful thing a pocket-map is!" I remarked. "That's another
> thing we've learned from your Nation," said Mein Herr, "map-
> making. But we've carried it much further than you. What do
> you consider the largest map that would be really useful?" "About
> six inches to the mile." "Only six inches!" exclaimed Mein Herr.
> "We very soon got to six yards to the mile. Then we tried a hun-
> dred yards to the mile. And then came the grandest idea of all! We
> actually made a map of the country, on the scale of a mile to the
> mile!" "Have you used it much?" I enquired. "It has never been
> spread out, yet," said Mein Herr: "the farmers objected: they said
> it would cover the whole country, and shut out the sunlight! So we

> now use the country itself, as its own map, and I assure you it does
> nearly as well."

1.2   Specific aims

Obviously, to be useful, a network model must be a simplified version of the
pattern of interactions in the complex system it represents, in the same way as
a good map is a simplified version of the landscape it represents. The challenge
is to create this map of some aspects other than the land that we stand upon. It
must be a dynamic map, because neither solar flares nor communicating people
are static systems. This calls for a map that is a full model with interacting
players over an ever-changing network—a dynamic network model. The main
goal with this thesis is to create a full network model like that. By characterizing
the world in this way I will not only be able to answer specific questions, but
also focus on asking general questions. For example, on the way to the dynamic
network model, it has been necessary to use, and sometimes develop, structural
and statistical measures that are informative not only for a particular network,
but for a spectrum of networks from biology to the man-made Internet.

To uncover the interplay between function and structure and the strong con-
nection to the formation process in networks, I have found it necessary to study
networks from a perspective of information transfer. In other words, I will treat
networks as information machines. With information I here mean any knowl-
edge that has been gathered or received by communication. The connection be-
tween information, communication, and networks is central in this thesis. I am
in particular interested in how the network structure affects the communication
conditions over the network, and vice versa, how the ongoing communication
affects the network structure. For example, who communicates with whom and
the social structure of a society are strongly integrated. The social network re-
flects the access to information that different parts of the system experience, and
social mobility may be seen as a quest for better information access.

The specific aim of this thesis is to introduce a model of self-organization
between communication and structure in networks, with a feedback between
different communication habits and the structure.

Moreover, maybe it is possible to find out how Shakespeare would have
formulated Hamlet's question with or without a connection to his contemporary
Marlowe, or in a network context other than his sixteenth century London?

1.3   Outline

There are already more than a handful good reviews on network science, and I would like to recommend the reviews by Boccaletti *et al.* (2006); Newman (2003b, 2005), the books by Newman *et al.* (2006); Bornholdt and Schuster (2003), and the Ph.D. thesis by Holme (2004b).

Instead of a full review, this thesis is first of all a journey that connects the articles that I have written together with my colleagues. The purpose is to present the ideas in a simple way, emphasize their interconnectedness and to create a synergetic picture. Several models and concepts are also visualized on http://cmol.nbi.dk/models.

The field of network science can be roughly categorized into three areas: empirical studies, network generating models, and dynamics on networks. This is also how the thesis is organized. It starts with a short overview to present the field of network science (Chapter 2). The overview, which also contains some important concepts and definitions that make the reading of the remaining thesis easier, is followed by the most common network models (Chapter 3). The central part of this thesis are Chapters 4 and 5, where the information/communication approach to network science is most evident. The last chapter in the first part of the thesis (Chapter 6) summarizes briefly the main findings of the thesis.

The second part of the thesis is the eleven papers the thesis is based on, listed on page v and hereafter referred to as Paper I, Paper II, and so on.

Chapter 2

Complex networks

Physics is a powerful science. Nevertheless, the deterministic Newtonian science lacks words to describe complex behavior. Physics has been the tool that break things down into laws. This reductionist viewpoint has proved to be very successful, but it is now time to reassemble things into a whole again (Wilson, 1998). There is much to be learned on this journey, and a new science has to be developed.

In the thesis I take the approach that complex systems are best understood through simple models. Simple, because it should be possible to understand the model in itself. Moreover, a model that contains everything that is known about the system can, by some fiddling with the parameters, in principle predict anything. And a model that can predict anything predicts nothing (Popper, 1963).

2.1   Background

Networks are everywhere, have always been, and will always be—social networks, gene networks, cellular networks, ecological networks, computer networks, phone-call networks, airline networks, power-line networks—why this buzz now? The short answer is the Internet.

A longer answer is that mathematicians have had, since Leonard Euler solved the so called *Königsberg bridge problem*[1] in 1736, the knowledge to abstract away

---

[1] Seven bridges connected the four different land masses in Königsberg, and the question was, "Does there exist any single path that crosses all seven bridges exactly once each?"

details from a problem and reduce it to a *graph*[2]—a set of nodes and links. The strength with this mathematical object is of course that the nodes are not limited to represent land masses and the links limited to represent bridges. The list of networks above could have been made much longer, since so many systems can be simplified with maintained complexity by a network.

It matters to abstract a system into its underlying network when the interacting units are *unique*, like humans, proteins, and web pages, but not when they are *interchangeable*, like atoms, electrons, and spins[3]. It is therefore not surprising that sociologists, and not physicists, were the first to apply the abstract mathematical objects of graphs to concrete problems. This happend around the 1950s and one challenging question was "How small is the world?" (Milgram, 1967; de Sola Pool and Kochen, 1978; Kochen, 1989). At about the same time the two Hungarian mathematicians Paul Erdős and Alfréd Rényi started to develop a statistical theory of dynamic networks (Erdős and Rényi, 1959), originally proposed by Solomonoff and Rapoport (1951).

Why did the field of network research wait another 50 years to take off? The main problem the sociologists faced was the difficulty to create large reliable networks. Especially de Sola Pool and Kochen (1978) asked many fundamental questions, still important today, but they could not answer them in an adequate manner because of the poor experimental data. For example, to estimate how long the shortest chain of friends is between two people chosen at random, they first had to estimate how many friends an individual has. The two problems they faced were: what is a friend and why are people so poor at estimating their number of friends? Try to estimate your number of friends and you will realize that it is not straightforward.

Kochen and de Sola Pool instead turned to the mathematical models by Solomonoff and Rapoport (1951) and Erdős and Rényi (1959), and estimated the average separation to three. Milgram (1967) estimated, by tracing letters sent by chains of aquaintaces, the average distance to six.[4] Obviously there was a problem here. The experiments predicted twice as large average separation as the models. Were the models wrong, or the experiments unsound?

One exception from small and unreliable data from experiments was presented in the pioneering work by Price (1965) on the network of citations be-

---

[2]The difference between the mathematicians *graph* and the *networks* discussed in this thesis is subtle, but where a graph is mainly a mathematical object, a network often refers to a decentralized, naturally evolving system (Newman *et al.*, 2006).

[3]Spin is the intrinsic angular momentum in elementary particles.

[4]The experiment will be thoroughly described in Chapter 3 on page 22.

tween scientific articles. He found that the distribution of citations, corresponding to links in network language, was not peaked at an average value as in the mathematical models of random networks (Erdős and Rényi, 1959). Instead most of the articles had only a few citations and a few articles had many citations: the distribution of citations per article followed a power law. A decade later he presented the model called *cumulative advantage* based on an idea by Simon (1955) that could explain this observation (Price, 1976).[5] However, the work by Price considered one particular network, and the concepts did not cross over to other sciences but remained an isolated work.

In the late 1990s, at the same time as the Internet started to become mature and useable[6], all necessary ingredients were suddenly at hand (Barabási and Albert, 2002): the emergence of large databases on the topology of various real networks, increased computing power, the noticeable breakdown of boundaries between disciplines, and the new holistic viewpoint to try to understand the system as a whole. Again it became apparent that it is impossible to overestimate the importance of cross fertilization between theory and experiments in science.

The breakthrough originated in the physics community because the tools of statistical mechanics are particularly well suited to analyze the patterns of interactions in networks. For example, in the context of statistical mechanics, physicists have investigated how a system behaves on the macroscopic level from the properties of its constituents, the *agents*, on the microscopic level. The new idea was that behind every complex system there is an underlying network with a nonrandom topology (Barabási, 2001). It became clear that this pattern of interactions, which forms the network, plays a fundamental role in understanding the system (Newman, 2002b).

2.1.1 Real-world networks

The first network that was large enough for reliable statistical measurements, and at the same time accessible, turned out to be the World Wide Web (WWW).[7] With the purpose to test a real-world network against the random network model (Erdős and Rényi, 1959), Barabási *et al.* (2000) analyzed a subdomain of the WWW consisting of more than 300,000 web pages. As in the study by

---

[5]This model is very similar to the preferential attachment model discussed in the next chapter.

[6]Notable: Ines Usman, the former Swedish secretary of transport and communications, said in 1996 that "The Internet is a fad that may just pass by."

[7]The network of the WWW consists of web pages linked by the hyperlinks we click on to go from page to page.

Price (1965), the distribution of the number of links per node did not follow the predicted distribution from the random network model. Again it was power-law distributed. However, this study did not remain isolated, because a number of experiments followed shortly. Broder *et al.* (2000) repeated the experiment on a much larger section of the WWW and confirmed the results. Faloutsos *et al.* (1999) and Vázquez *et al.* (2002) showed that the distribution of physical connections in the Internet followed roughly the same distribution.

Were the discovered power-law tails of the distributions of links "much ado about nothing?" Yes and no. Yes, because the discovery has led to an unreasonable hunt for power laws in empirical studies and the construction of ad hoc models that produce power laws. No, because the few nodes with a very high number of links, the hubs of the network, have a dramatic effect on the properties of the network.[8] Moreover, the broad distribution was discovered in other networks than technological as well. Watts and Strogatz (1998) studied movie actor collaborations, Fell and Wagner (2000) the metabolic network of the bacterium *E. coli*, Jeong *et al.* (2000) the metabolic networks of more than 40 organisms, Amaral *et al.* (2000) airline networks, Liljeros *et al.* (2001) sexual contacts, and we have for example studied city street networks (Paper V). The list can be made much longer, but the general result was that the networks did neither look like the random network model predicted, nor were they regular, and what they really look like is what the remaining of this thesis is about.

2.2   Network topology

I will begin this section with two examples of real-world networks. This will not only visualize the pattern of links connecting nodes, the *network topology* or the *network structure*, in two different systems. It will also illustrate the difference between a cumbersome and detailed, and an efficient but inexact way to extract the underlying network in a system.

2.2.1   Two network examples

The first network in Figure 2.1 is by Krebs (2002) and describes the September 11 terrorist network. With the aim to "uncover network patterns that would reveal Al Qaeda's preferred methods of stealth organization," Krebs gathered in-

---

[8]Pastor-Satorras and Vespignani (2001) showed that networks with power-law distribution of links per node may not have an epidemic threshold and a disease will therefore always spread.

formation from major newspapers such as the *New York Times* and the *Wall Street Journal*. Obviously this method is time consuming and inevitably incomplete.

The second network in Figure 2.2 consists of a subset of the 100 most influential people in 2006 selected by *Time Magazine*. In this case I have let *Google* create the network by asking how many hits two names get together compared with what one of the names gets. If the ratio is above a threshold, 5% in this case, I create a directed link between the two names pointing to the name I compare with. This is a very fast method, but gives of course only a proxy of the real network.

I have already stated that real-world networks are different from the random network model by Erdős and Rényi (1959). This difference refers to the ways the links are organized between the nodes—the network topology. It is for example easy to see that the links in the two networks in Figures 2.1 and 2.2 are not randomly distributed, but how much, and in what way are they not? Before answering these questions, we need a benchmark to enable comparison between networks through proper null models. However, I have already used too many concepts that deserve to be defined rigorously before I use them, so it is time to take a step back.

2.2.2   Nodes and links

Technically speaking, a mathematician uses the term *graph* to represent a structure with a set of objects called *vertices* connected by links called *edges*. If the edges have directions (object A links to B, but B does not automatically link to A), the graph is directed. If the edges are also associated with a weight, that for example could represent the distance between two objects, the mathematician uses the term *network*. However, that *network* and not *graph* will be used from now on does not imply that this thesis is about directed weighted graphs. Instead, undirected and unweighted graphs are almost exclusively the focus of this thesis. This contradiction originates from the loose language outside mathematics and the tradition to use the term network as something that represents a real-world evolving system upon which distributed dynamic processes might exist (Newman *et al.*, 2006). I will also use the terms *nodes* and *links* instead of vertices and edges.

The simplest concepts beyond the nodes and links, still belonging to the local properties of a network, are the *degree* and the *clustering*. They will be explained in the following sections together with the *shortest path*. Degree, clustering, and shortest path were three major concepts in focus in the early studies

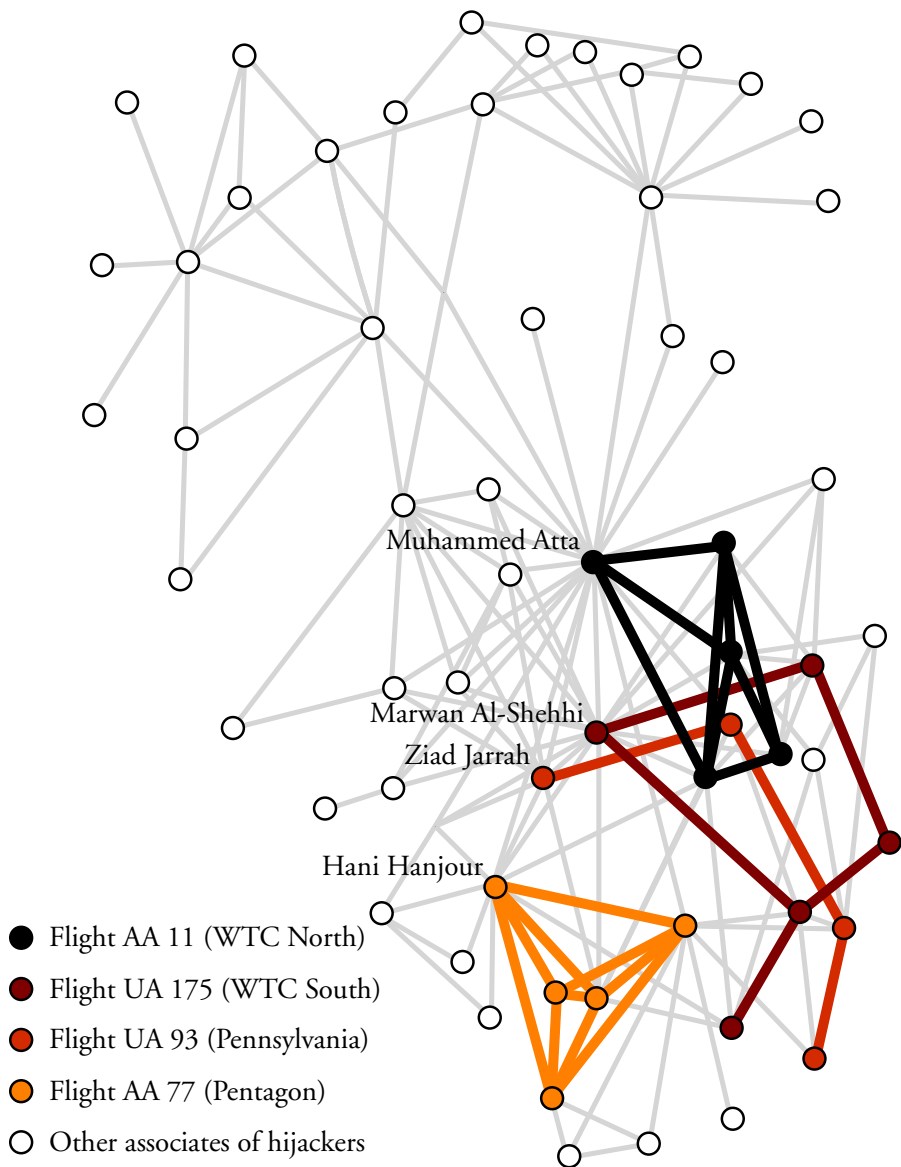Figure 2.1: *Terrorist network of the September 11 hijackers and associated (Krebs, 2002). In total 62 terrorists, with the names of the four pilots. The terrorists on each plane are identified by separate colors.*
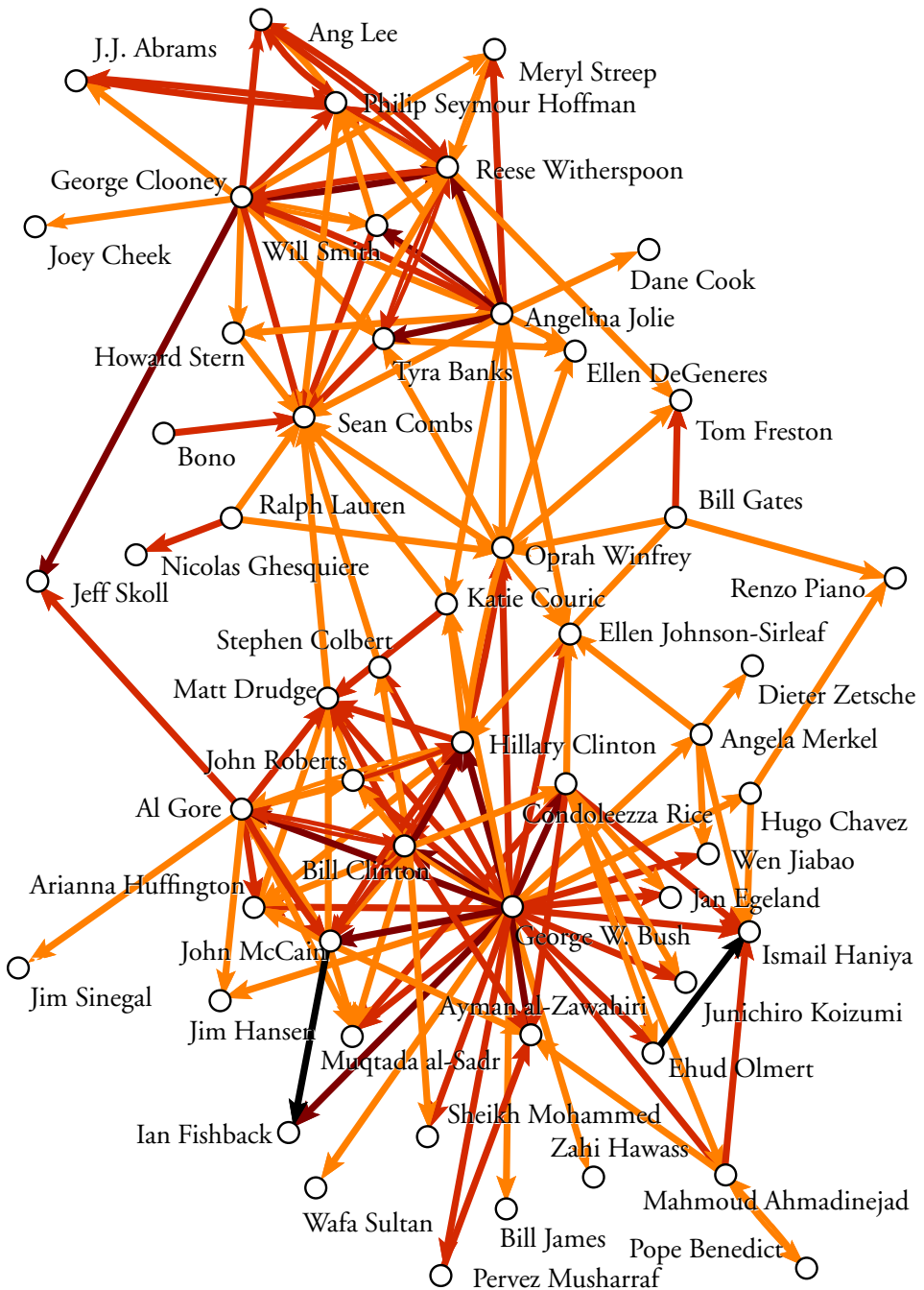
Figure 2.2: *A subset (54) of the 100 most influential people in the world selected by Time magazine, and connected by Google in June 2006. A strong connection is represented by a dark link.*

of complex networks, not least because of their relevance to the fundamental questions raised by de Sola Pool and Kochen (1978) (see page 6).

2.2.3   Degree distribution

The *degree k_v* is the number of links incident to node v and is consequently equal to the number of nearest neighbors of v. In physical literature the degree of a node is also known as the node's connectivity. There is a spread in the node degree, described by the *degree distribution*, which is characterized by the *distribution function $P(k)$*. $P(k)$ gives the probability that a randomly chosen node in the network has exactly k links or equivalently k nearest neighbors. The degree distribution plays a central role in Chapter 3. Power-law degree distributions have already been mentioned, and refer to distributions that are proportional to $P(k) \propto k^{-\gamma}$. A real network does not of course exactly follow a given parametrization of a distribution, but if it is roughly $P(k) \propto k^{-\gamma}$ it will be called a broad degree distribution in this thesis. The important feature in a network with a broad degree distribution is that most nodes have very few links, but a few nodes have very many links. The nodes with very many links are called *hubs*.
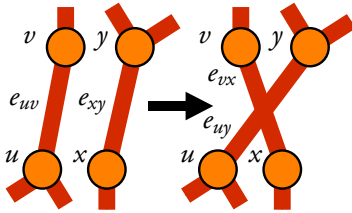
2.2.4   Degree correlations

Figure 2.3: *The elementary step in the random rewiring introduced by Maslov and Sneppen (2002a). The links e_{uv} and e_{xy} are chosen at random and rewired to become e_{uy} and e_{vx} provided that none of these links already existed.*

Maslov and Sneppen (2002a) developed a method to test for correlations in the degree of the nodes. The probability $P(k_0, k_1)$ that two nodes of degree k_0 and k_1 are connected via a link is calculated and compared with the probability $P_r(k_0, k_1)$ in a randomized version of the same network with preserved degree sequence. The idea is a reinvention of the conditional uniform test by Katz and Powell (1957), presented in a network context by Snijders (1991).

In the *randomized counterpart* of the original network, the *null network*, all nodes have the same degree as in the original network, but the connections are

randomized. The random counterpart is created by first randomly selecting a pair of links, like e_{uv} and e_{xy} in Figure 2.3. The two links are then rewired in such a way that u becomes connected to y and v to x, or vice versa provided that none of these links already existed. The step is iterated until all information of the original network is lost.

If the above algorithm is repeated, a set of randomized counterparts makes average expectation values and standard-deviation measurements possible for any property of the network. The randomized networks will hereafter be called MS--randomized networks (from Maslov and Sneppen) to avoid confusion with the random networks, already mentioned but more thoroughly presented in Chapter 3. The ratio of a property between the original and MS-randomized versions uncovers deviations from random properties. The null-network approach is extremely powerful to separate nonrandom features from features that are only associated to a particular degree distribution and will be used over and over in this thesis.

Summary of paper I
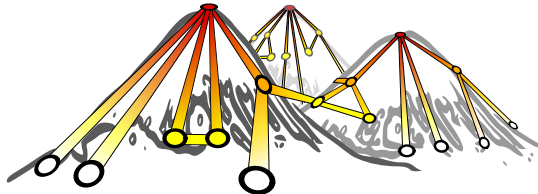
Figure 2.4: *A network as a degree landscape with mountains and valleys, with the altitude of a node proportional to its degree.*

In Paper I we generalized the degree-organizational view (Pastor-Satorras *et al.*, 2001) of real-world networks with broad degree distributions (Barabási and Albert, 1999) in a landscape analogy with mountains (high-degree nodes) and valleys (low-degree nodes) (see Figure 2.4). For example, with the presented viewpoint, correlated degrees between adjacent nodes correspond to smooth landscapes (social networks) (Newman, 2002a), hierarchical networks to one-mountain landscapes (the Internet) (Trusina *et al.*, 2004), and networks with anti-correlated degrees without hierarchical features to rough landscapes with several mountains (Newman, 2003a) (see Figure 2.5).

(a)   (b)   (c)

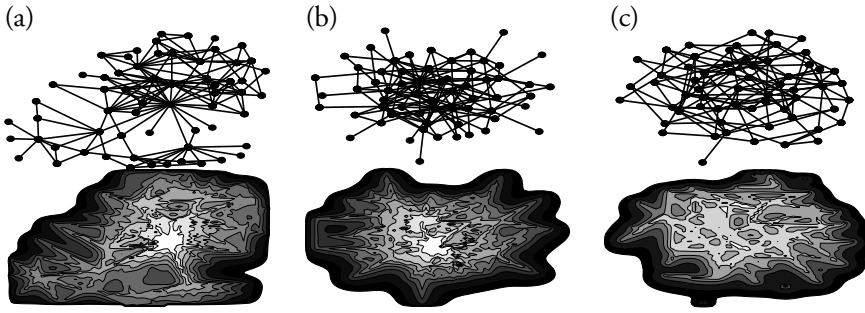Figure 2.5: *The terrorist network in Figure 2.1, here as degree landscapes. The landscape in (a) corresponds to the original network, in (b) it is randomized with fixed degree distribution and in (c) it is completely reshuffled without any constraints except for the number of nodes and links. The important difference between (a) and (b) is that (b) essentially only contains one mountain, where it is a more heterogeneous landscape in (a). All high degree nodes have disappeared in the completely reshuffled network in (c). This is clear in the topological maps (bottom), where the mountain is washed out in (c).*

To quantify the topology, we measured the widths of the mountains and the separation between different mountains. We also generated ridged landscapes to model networks organized under constraints imposed by the space the networks are embedded in, associated to spatial or, in molecular networks, to functional localization.

2.2.5 Shortest path

The length of all links is here defined to be 1. The *distance* $l_{uv} = l_{vu}$ between two nodes u and v in the network is consequently defined to be the number of links traced on the *shortest path* between the nodes. A *connected component* is a set of nodes that all can be reached from each other, disconnected from other components of the network. If the network only contains one connected component, the network is *connected*. When this is true, it is straightforward to define the *average path length l*, as the average over all shortest paths

$$l = \frac{2}{N(N-1)} \sum_{u \neq v} l_{uv}. \tag{2.1}$$

The node that on average has shortest path length to all other nodes in the network can naturally be defined to be the *center node*.

If the network is disconnected the *average reciprocal distance l^{-1}*, also called the *efficiency*, can be used instead of the average path length (Latora and Marchiori, 2001)

$$l^{-1} = \frac{2}{N(N-1)} \sum_{u \neq v} \frac{1}{l_{uv}}. \tag{2.2}$$

A measurement related to the shortest path is the *betweenness*. The betweenness of a node is the number of times the node is on the shortest path between all other pairs of nodes. The quantity was introduced to investigate the participation rates of nodes in networks with the shortest paths as proxies for communication channels (Freeman, 1977). An alternative center node in this context is the one with the highest betweenness.

2.2.6 Clustering

The *clustering coefficient C* is the probability that the two nearest neighbors of a node also are nearest neighbors to one another. In the network, such a relation forms a *triangle* of links. Accordingly, a high *clustering coefficient* reflects an increased number of triangles in the network, and gives a hint about a dense topology on the local network level. A common definition of the clustering coefficient is given by Barrat and Weigt (2000):

$$C = 3 \frac{N_\triangle}{N_\sqcup}, \tag{2.3}$$

where $N_\triangle$ is the number of *triangles* in the network and $N_\sqcup$ is the number of *connected triples* of nodes in the network. Triangles are trios of nodes with every node connected to both of the others. Connected triples are trios of nodes with at least one connected to both of the others. The factor 3 compensates for the fact that each triangle corresponds to three connected triples. The factor constrains C to lie in the range between 0 and 1. In a social network, where nodes are persons and links symbolize friendship, C corresponds to the probability that two of one's friends also are friends with one another (remember the questions raised by Kochen that I mentioned on page 6). Whereas the shortest path is a global property of the network, clustering is a local property.

The drawback of the definition of the clustering above is that it is dependent on the global properties of the network. To overcome that problem we use the null-network approach. We think that the best way to calculate the clustering

in a network is to count the number of triangles in the network and compare with the number of triangles in its random counterparts. This is a good way to avoid the direct effect of the degree distribution on the clustering, and we use this method in all presented papers.

2.3   Beyond simple measures

There are plenty of other measures invented to reveal the structure of real-world networks, too many to all be mentioned here. However, by just looking at the networks in Figures 2.1 and 2.2 it is clear that they both have a higher order structure than the broad distributions of links and the many triangles. It looks like the terrorist network could easily be divided into three groups, and the network of influential people could be divided into two groups. By looking more closely at the names it is clear that political leaders form one group and people from showbiz form the other group, and that Oprah Winfrey falls somewhere in between.

2.3.1   Community structure

No one has contributed more than Newman (2004a) on how to detect groups in networks, or *community structures*, which they are often called. For example, Girvan and Newman (2002) presented a method based on the betweenness. They showed that it is possible to partition the network by successively removing the link with the highest betweenness until the network falls apart into two disconnected components. This brute force method can then be repeated on the components formed.

Later Newman and Girvan (2004) introduced a measure of the community structure under the name *modularity*. It resembles the process we perform when we look at the networks in Figures 2.1 and 2.2 to identify the groups. A meaningful division of the networks is not only one in which there are few links between the groups—it is one where there are fewer than expected. They formulated this mathematically and have shown that it gives rise to meaningful partitions of networks (Newman, 2004b). Newman (2006) presented a method for the maximization of the modularity over possible divisions of a network based on a matrix method that is very efficient.

The other way to look at the problem is to study the formation process, and what dynamic features the community structure brings? This is indirectly the topic of Chapters 4 and 5. The formation process has also been studied

by for example Grönlund and Holme (2004), and the effect of the community structure on the spread of fads by Grönlund and Holme (2005). For example, Boguñá and Pastor-Satorras (2002) showed that the epidemic threshold can be nonzero even in networks with power-law degree distribution if the network is modular or has high clustering.

Before going into more detail here, and to be able to connect the network topology with its intrinsic function and the underlying evolutionary process, the next chapter will be about network models.

Chapter 3

Network models

Constructing a network model is a way to understand the formation process of the network, and thereby also shed light on the function of the network. A network model is nothing but a set of rules that in a static or dynamic way in the end specifies what connects to what. A network model can also create generic classes of networks, which can be used as playgrounds for dynamic models on networks. For example, to thoroughly examine vaccination strategies on different network topologies.

All network models are originally promoted by empirical studies. Observed topological features, such as short path-length, high clustering, and nonrandom degree distributions, have been followed by models that reproduce the desired attributes in an attempt to uncover their dynamic origin (Watts and Strogatz, 1998; Barabási and Albert, 1999; Klemm and Eguíluz, 2002). We learned from the quotation of *Sylvie and Bruno Concluded* by Lewis Carroll (page 2) that a good model is a simple model, and details are therefore by necessity left out. This is the essence of the different classes of network models presented in this chapter.

3.1 Simple networks

A lattice is the simplest possible interaction pattern (Figure 3.1(a)). The extent to which it has been used in statistical physics, to model dynamic systems, can not be overestimated. In fact, many systems can be accurately modeled by having their parts interacting on a lattice layout. For example, magnets are accurately modeled by interactions on a lattice in the Ising model (Ising, 1925), simply
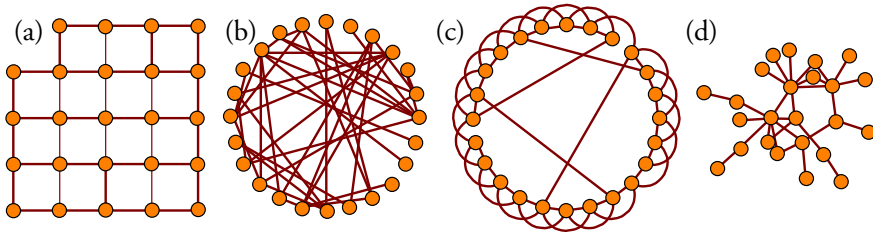
Figure 3.1: *Four different network models, all designed to mimic real-world networks. (a) A 2-dimensional lattice. (b) n = 24 nodes and m = 60 links corresponding to p = 0.14 in the random network model by Erdős and Rényi. The average degree is consequently ⟨k⟩ = 4. (c) The small-world model by Watts and Strogatz; 24 nodes, each connected to the first and second nearest neighbor in the one-dimensional ring. 4 links are rewired, which corresponds to a probability of rewiring p = 0.08. (d) A scale-free network with degree distribution $P(k) \propto k^{-2.2}$.*

because the lattice is a good representation of the real interaction pattern. A lattice is also a good choice if the scope is to understand the intrinsic features of a model. The results can in some cases be derived analytically, since the effect of the interaction pattern is the simplest possible. However, a lattice is not a good network model when the effect of the interaction pattern itself plays an important role, since it is a bad representation of typical interactions in the real-world (Watts and Strogatz, 1998).

3.2   Random networks

Random networks, or ER-networks after Erdős and Rényi (1959), briefly discussed in Chapter 2 and illustrated in Figure 3.1(b), should not to be mixed up with the MS-randomized networks in Figure 2.3. ER-networks play a central role in the study of complex networks because the underlying mechanisms of organization in nature is sometimes random (Barabási and Albert, 2002). The advantage with the ER-model presented below is, thanks to its simplicity, that many properties can be analyzed analytically.

(i) The total number, n, of nodes is fixed;

(ii) The probability that two arbitrary nodes are connected is p.

The average number of links is consequently

$$\langle m \rangle = pn(n-1)/2. \tag{3.1}$$

The degree distribution is binomial

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}, \tag{3.2}$$

since for a certain node the probability of k links is p^k, the probability of absence of extra links is $(1-p)^{n-1-k}$ and there are $\binom{n-1}{k}$ possible end nodes for the k links. For large n, $P(k)$ takes the form

$$P(k) = e^{-\langle k \rangle} \langle k \rangle^k /k!, \tag{3.3}$$

which is the narrow Poisson distribution, peaked at the average degree

$$\langle k \rangle = p(n-1) \approx pn. \tag{3.4}$$

Random networks have been extensively studied and one conclusion is that for $\langle k \rangle \gtrsim \ln(n)$ almost every network is connected and the average shortest path is (Dorogovtsev and Mendes, 2002; Barabási and Albert, 2002)

$$l_{rand} \sim \ln n / \ln \langle k \rangle. \tag{3.5}$$

The clustering coefficient is

$$C_{rand} = p \approx \frac{\langle k \rangle}{n}, \tag{3.6}$$

since the probability that two of a node's neighbors are connected is p.

To study feature deviations in a network from an ER-network, *random ER-counterparts* of the network can be generated in a similar way as the MS-randomized networks. A link is chosen randomly together with two unique nodes that become the new endpoints of the rewired link. This procedure washes away all nonrandom properties, also associated to the degree distribution. However, we seldom use this strategy since it has the same disadvantage as the definition of the clustering coefficient in Equation (2.3), in that it disregards the extensive effects the degree distribution itself has on many properties.

The ER-networks share the short average path length, Equation (3.5), with networks in nature, but for large networks the clustering coefficient, Equation (3.6), is negligible. This is an unrealistic property, since networks in nature are characterized by a high clustering coefficient.

3.3   Small-world networks

The concept of a small world dates from the famous experiment by Milgram (1967). The social psychologist Stanley Milgram sent letters to a few hundred randomly selected individuals in Nebraska and Kansas, USA, with the instruction to the receiver to reach one of two target persons in the Boston area (Milgram, 1967; Kleinberg, 2000b). The recipients were told to send the letter to a person they knew on first-name basis, an acquaintance, who they thought was more likely to know the target than they were themselves. Subsequent recipients followed the same procedure. The information given about the target was their name, address, and occupation. The letters were traced by requesting that each participant should send Milgram a postcard. The result was that the average length of the resulting acquaintance chains was about six. This is popular known as the principle of six degrees of separation and was the first evidence that the world is indeed small (Watts, 2003).[1]

Random networks have the small world property, due to the typical node–node distance $l \sim \ln n / \ln \langle k \rangle$. However, the clustering coefficient $C = \langle k \rangle / n$ is much smaller than for most real-world networks (Newman, 2002b). One of the first models with both the property of high clustering coefficient and the small-world effect was the small-world model of Watts and Strogatz (1998), also known as the WS-model. Ball *et al.* (1997) presented at the same time a very similar model in the context of spreading of epidemics.

In the WS-model, the network is created from a one-dimensional lattice with periodic boundary conditions (see Figure 3.1(c)). Every node is not only connected to the nearest neighbor, but also to at least one next-nearest neighbor in the one-dimensional lattice. In this way, a node has four or more nearest neighbors and not only two, which would result in a vanishing clustering coefficient. Then, one by one, every link is rewired with probability p in such a way that long range connections appear. In other words, many links connect locally and a few links are used for long range connections as in Figure 3.1(c). The model is therefore a way to go from lattice models to random networks by adjusting the parameter p. For large p the WS-model is the same as the random network. But even for small probabilities of rewiring, where the local properties are almost unchanged, the average shortest path length is of the order of that for random networks. The short-cut links might represent acquaintances one

---

[1]Although this was the first real evidence that the world is small, the concept was not new. Frigyes Karinthy published 1929 *Chains*, where he argued that everyone on earth is at most five acquaintances away from anyone else (Newman *et al.*, 2006).

has in other states or countries, explaining the short acquaintance chains in the experiment by Milgram.

The WS-model was one of the first models that could simultaneously explain high clustering and small-world effects. However, the discovery of broad degree distributions in real-world networks ignited a burst of new models to explain this feature (Faloutsos *et al.*, 1999; Barabási and Albert, 1999; Caldarelli *et al.*, 2002; Goh *et al.*, 2001).

3.4   Scale-free networks

Today information about different networks in nature is becoming increasingly available (Barabási *et al.*, 1999). The small-world effect and high clustering in real world networks, which have long been assumed to exist, have now been validated. Much more unpredicted was the recent discovery that the degree distribution of at least several large (growing) networks in nature followed a power law (Faloutsos *et al.*, 1999; Barabási and Albert, 1999) (Figure 3.1(d)). As mentioned before, the power-law distribution has been reported for the Internet (Faloutsos *et al.*, 1999), the WWW (Albert *et al.*, 1999), some molecular networks (Jeong *et al.*, 2000), as well as for the size distributions of industrial companies (Zajdenweber, 1995) and much earlier for citation networks (Price, 1965). However, the *Zipf law* has been known for a long time in a nonnetwork context, for example that the rank of word frequencies, city sizes and income distributions follow a power law (Zipf, 1949).

The power-law distribution of degrees k is given by

$$P(k) \sim k^{-\gamma}, \tag{3.7}$$

where γ is approximately between 2 and 3 in real-world networks (Dorogovtsev and Mendes, 2002). Barabási and Albert (1999) called such networks *scale free*, since they are free of a characteristic scale—they are scale invariant. In other words, the average degree $\langle k \rangle$ is an insignificant quantity in a power-law distribution. Contrary, a node in a network with an exponential or Poisson degree-distribution typically has degree $k \sim \langle k \rangle$. The main feature of scale-free networks, is that each node has a statistically significant probability of having a very high degree compared with the average degree $\langle k \rangle$ of the network. These highly connected nodes dominate the topology of the network by forming hubs, which is not the feature of random or small-world networks (see Figure 3.1).

3.4.1   The Barabási-Albert Model

Barabási and Albert (1999) argued that the scale-free nature of these networks originated from two generic features of many real-world networks: First, most real networks are open systems that *grow* by the addition of nodes. Second, the nodes are not linked randomly. There is a *preferential attachment*, which favors attachment to nodes with a high degree. This manifests itself in facts like "rich gets richer" and that being popular is attractive. As was mentioned in Chapter 2, the idea was originally modeled in a nonnetwork version by Simon (1955) and in a network version by Price (1976) to explain the power-law degree distributions of citation networks (Price, 1965). However, I will here present the Barabási-Albert (BA)-model (Barabási and Albert, 2002), which contains three elements:

> *Initial condition:* The network consists of a small number n_0 of nodes and no links.

> *Growth:* At every time step, a new node with $m < n_0$ links is added, linking to m different nodes already present in the system.

> *Preferential attachment:* The probability P that the new node will be connected to node i depends linearly on the degree k_i.

The power-law exponent γ is 3 in this model, independent of m (Barabási and Albert, 1999).

The continued attempts to find better models are driven by the discovery of new topology properties of real-world networks. For example, the model introduces correlations among the nodes (Grönlund *et al.*, 2005) and other unnatural features, such as a vanishing clustering coefficient as $n \to \infty$ (Barabási and Albert, 2002). The fact that the power-law degree distribution only arises for linear attachment (Krapivsky and Redner, 2002), questions the model's generality. Another unappealing feature of the BA-model is that the networks are in some way dead, as it is only a slight generalization of Simon's nonnetwork model in a network context. This means that if the model is a good description of the evolution of networks in nature, then the question "Who is connected to whom?" looses its value, since only the degree of the nodes counts.

Also, many networks in nature with a broad degree distribution show deviations from a pure power-law, typically an exponential cutoff at high degrees

$$P(k) \sim k^{-\gamma}\phi(k/\xi), \tag{3.8}$$

where $\phi(k/\xi)$ is the cutoff at some scale ξ. In the context of the growing BA-model, this phenomenon can be explained if an aging of nodes is introduced (Barabási and Albert, 2002). The effect is that old nodes that are added early in the network-evolution process, lose some of their attractive power. This limits the ability for nodes with a very high degree to evolve.

Despite my criticism, I accept without question the necessary impact on the network community from the preferential-attachment model. In particular, attention has shifted to growing networks to capture the features of real-world networks since the BA-model was introduced.

The BA-model belongs, together with the ER-model and the WS-model, to the first generation of simple models. However, local and global optimization models inspired by evolutionary processes, have been developed in the field of biology (West *et al.*, 1997). Local optimization models of nongrowing networks have, on the other hand, scarcely been investigated. Two classes of models of this kind will be presented in this thesis and merging is the first.

3.4.2 Merging

Information transfer is the fundamental task of many networks in nature. In these networks, some of the nodes receive signals from outside and then transfer these signal to other nodes, which then take appropriate action. Information networks include for example neural networks with their synaptic rewiring (Cohen-Cory, 2002), molecular networks evolved to modulate the protein production in living cells (Jeong *et al.*, 2000; Maslov and Sneppen, 2002a), and social networks exemplified by the Internet (Faloutsos *et al.*, 1999). Networks where material or energy is transmitted between nodes, as for example business networks or ecosystems, often also contain an inherent ingredient of information transfer. Therefore, a natural source of inspiration to invent new models is to consider real-world networks in the perspective of information transfer.

Every action or process in and between the subunits of a system, here simplified and represented by nodes, is more or less complicated and time consuming. Is there a way to optimize the system, locally or globally, so that the overall functionality increases? Two appealing modifications of the network would be better access between nodes (shorter shortest paths) and more efficient and possibly larger computing units. This inspired Kim *et al.* (2005) to introduce the idea of *merging-and-creation* on networks, exemplified in Figure 3.2 by the net-
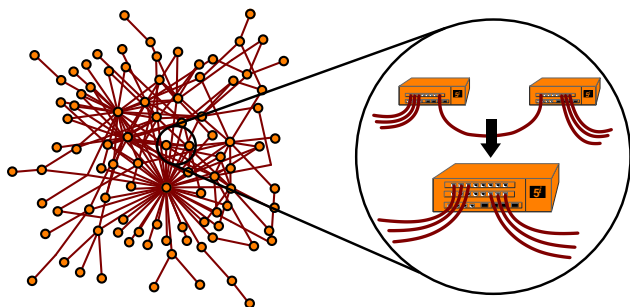
Figure 3.2: *The core in the merging model: To the right two computing units in a network merge to a larger unit. The reason could be economical, based on efficiency or performance. This is a likely process in for example the network of autonomous systems of the Internet (modeled to the left by a scale-free network with $P(k) \propto k^{-2.2}$), where companies that own the autonomous systems merge, mainly for financial reasons.*

work of *autonomous systems*[2] of the Internet.[3] One version of the merging and regeneration model is:

- Choose a node i and one of its neighbors j randomly.

- Let i and j merge by absorbing the link between them. Choose one of the nodes as the surviving node (i). By copying all links from j to i that i did not already have (all double links are removed), i has now the same neighbors as i and j had before the merging event.

- Add a new node with k_{new} links to k_{new} nodes in the network. The distribution of k_{new} is not essential for the model.

The regeneration step exists to keep the network alive with a fixed number of nodes. The model presented above gives rise to a power-law degree distribution $P(k) \propto k^{-\gamma}$ with a slope γ between 1.5 and 3 depending on variations of the model.

Summary of paper II

In Paper II we investigated different variations of the model and primarily the effect on the degree distribution. We also considered a nonnetwork implemen-

---

[2]An autonomous system is a collection of Internet-protocoll networks and routers with a common routing policy under the control of typically an Internet service provider.

[3]A simulation is available on http://cmol.ni.dk/javaapp.php

tation of the model, to be able to treat the model analytically. Alava and Doro-govtsev (2005) considered later a simplified network implementation analytically. Instead of merging nodes we let *elements* q_i without any links merge, in a large system ($i = 1, 2, \ldots, n$). The scalar may be either just positive (Field and Saslow, 1965) (Figure 3.3(b)) or both positive and negative (Takayasu, 1989; Krapivsky, 1993) (Figure 3.3(a)). To get a real-world idea about the simplified model, consider the elements as particles and the scalar as the mass of a particle in a model of dust aggregation in space (Field and Saslow, 1965). Other examples are companies and the financial assets of the company, and the vorticity of vortices studied in the context of superconductors and turbulence.
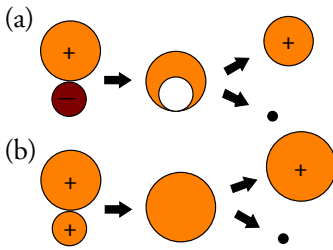

Figure 3.3: *One realization of the merging algorithm with different sign in (a) and equal sign in (b).*

In the core of the process two randomly chosen elements, i and j, merge and the new element acquires the sum of the scalars $q_i + q_j$. In parallel, there is a creation process of elements with small size $|q|$, both positive and negative. I emphasize that the size of the created elements is not essential for the dynamics, since the size of a new element can be drawn from a narrow distribution or be assigned deterministic values. Obviously there are many variants of this process, but they can be classified into three main classes according to the scaling of the size distribution $P(q) \propto q^{-\gamma}$ of the elements. (see Table 3.1).

Table 3.1: *Three different classes of the merging. In the first model all elements $q \geq 0$, but in the second model the elements q can be both positive or negative. In the third model a condition for the merging-and-creation event is that $q \geq 0$ and r is a random number from a narrow distribution $\langle r \rangle = 1$*

γ	3/2	2	5/2
merging	$q_i \rightarrow q_i + q_j$	$q_i \rightarrow q_i + q_j$	$q_i \rightarrow q_i + q_j - 1$
creation	$q_j \rightarrow 1$	$q_j \rightarrow \pm 1$	$q_j \rightarrow r \in [0, 2]$

Summary of paper III

In Paper III we studied one of the models in detail, the $\gamma = 2$ version, and re-
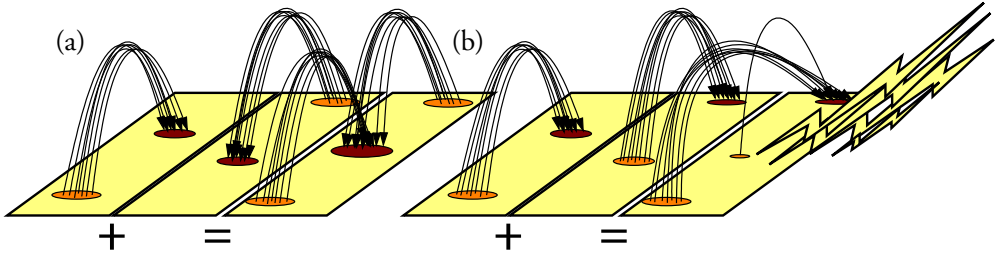
Figure 3.4: *Merging (a) and annihilation (b) in a model of the magnetic network of flux tubes on the sun. Positive nodes (donors) have outgoing links and negative nodes (acceptors) have incoming links. This $\pm$-symmetric dynamics give rise to a scale-free size distribution of both donors and acceptors ($P(q) \propto q^{-2}$) and energy bursts scaling as $P(s) \propto s^{-3}$, with s being the size of a burst.*

turned to a network implementation.[4] The process we studied is the completely $\pm$-symmetric case. A link has a positive end connected to a donor node (+ node) and a negative end connected to an acceptor ($-$ node). Interesting with this model is the close connection to the magnetic network of flux tubes on the sun and the associated size distribution of solar flares and sun spots (Lu and Russell, 1991; Hughes *et al.*, 2003) (see Figure 3.4). The model is:

- Merge the two random nodes i and j. There are now two possibilities:
  (a) If they have the same sign all the links from i and j are assigned to the merged node. Thereby the merged node has the same neighbors as i and j had together prior to the merging (Figure 3.4(a)).
  (b) If i and j have different signs, the resulting node is assigned the sign of the sum $q_i + q_j$. Thereby a number $max\{|q_i|, |q_j|\} - |q_i + q_j|$ of links are annihilated in such a way that only the two merging nodes change their number of links. This is done by reconnecting donor nodes of incoming links to acceptor nodes of outgoing links (Figure 3.4(b)).

- One new node is created of random sign, with one link being connected to a randomly chosen node.

This is a minimalistic model of the more complicated 2D-model presented by Hughes *et al.* (2003). Nevertheless, it gives the same qualitative results. The degree distribution is $P(q) \propto q^{-2}$ both for donors and acceptors and the size

---

[4]A simulation is available on http://cmol.nbi.dk/javaapp.php

distribution of the solar flares is $P(s) \propto s^{-3}$. With the degree of a node interpreted as the size of a magnetic flux line, the results agree with other modeling (Hughes and Paczuski, 2003; Lu and Russell, 1991). However, the size distribution of solar flares measured by Dennis (1985); Lin *et al.* (1984) follows a power law $P(s) \propto s^{-1.7 \pm 0.3}$ that disagrees with our result. According to Lu and Russell (1991), the explanation is that the annihilation of solar flares triggers an avalanche of events that makes the size distribution of the solar flares broader.

The advantage with our model is that it can be treated and understood analytically. Its simplistic form also opens for a coarse understanding of the fundamental mechanisms of some dynamic systems, not necessarily limited to sun processes in the magnetic corona. The total dominance of preferential attachment in the network community has overshadowed and possibly hindered other dynamic models to be investigated in detail. I believe that merging is one of the underestimated mechanisms that must be investigated further in the search for a complete understanding of the connection between the evolution of networks and their structure and function.

3.5 Beyond simple models

As already mentioned, there are numerous other models that produce power-law degree distributions. However, instead of giving examples I would like to mention a model that does not yet exist.

The *tolerance* is the inherent robustness of a network (Albert *et al.*, 2000). An *error* is defined as the successive removal of randomly chosen nodes and an *attack* as the removal of a few nodes that play a vital role in maintaining the network's connectivity. Vital nodes are those with the highest degree. Compared with random networks the tolerance against random failures, that is errors, is high in scale-free networks. The reason is that a high percentage of the nodes have just a few connections in a scale-free network. However, the price for a high error-tolerance is the vulnerability to attacks; when only a few of the dominating nodes are removed the network falls apart.

The question is therefore, what does the topology look like in the scale-free network that is optimized against attacks (Boccaletti *et al.*, 2006)? A better understanding is of crucial importance to ensure the safety, capacity, performance and to use all of the possibilities of man-made structures like the Internet and WWW (Dorogovtsev and Mendes, 2002).

Chapter 4

Navigation and the information horizon

L ife without information is not life. From the genetic blueprint in our DNA to the world-wide Internet, information and its dynamic counterpart communication are fundamental in our civilization. However, we live under the limited information horizon, in the sense that information is often imperfect and communication is always finite. In this chapter this horizon will be quantified.

4.1 Structure and function

The degree distribution alone is not enough to fully characterize a network. In order to reveal the function of networks and to understand the intrinsic mechanisms of an evolutionary process, higher order nonrandom patterns must be considered. One of the simplest such pattern is the degree-degree correlations presented in Chapter 2. The approach follows the strategy that we always use; measure a quantity on the network, randomize the network with conserved degree distribution, and measure again (Maslov and Sneppen, 2002a). If the randomization is repeated the statistics can be used to estimate the significance of the difference (Maslov and Sneppen, 2002b,a). If there is a difference our interpretation is that there is a nonrandom feature in the network. The feature can be a result of the evolutionary process and has probably a meaning for the global and local function of the network. A good example of a higher order topology investigation is the degree hierarchy presented by Trusina *et al.* (2004). Based

on the assumption that shortest paths between nodes can be used as a good approximation for communication pathways, Trusina *et al.* (2004) investigated the characteristics of such paths with respect to the degree sequence along the path and asked the question "To what extent are the shortest paths hierarchical?" In detail, what is the fraction of shortest paths that first visit nodes with higher degrees than the source node, reach a maximum value, and then visit nodes with gradually decreasing degrees before they reach the target node. A well defined measurement, but nevertheless blunt in itself. For a straightforward result, the investigation must contain a comparison with a null network, a randomized counterpart of the investigated network. In this way Trusina *et al.* (2004) managed to qualitatively and quantitatively see completely different communication patterns between the man-made network of the Internet and the biological network of yeast, which were not revealed by the degree distribution alone.

As each element interacts directly only with a few particular elements in most complex systems, distant parts of the network must consequently communicate through sequences of local interactions. In this way all parts of the network can be reached from other parts, but not all such communications are equally easy or accurate (Friedkin, 1983; Kochen, 1989; Milgram, 1967). In the folowing sections I will present what we believe is a good way to characterize the interplay between searchability of a network and the network structure. By searchability or navigability we mean the difficulty of sending a signal between two nodes in a network without disturbing the remaining network. This approach is therefore different from the extensive literature considering spreading of epidemics and rumours on networks (Balthrop *et al.*, 2004; Newman *et al.*, 2002; Newman, 2003b; Holme, 2004a; Watts, 2002).

4.2   Hide and seek on complex networks

As many networks are representations of the communicating backbone between elements in a system, surprisingly little effort has been put into qualitatively and quantitatively establishing the connection between the ability to communicate in a network and the network topology. The shortest path presented in Chapter 2 is the simplest approximation of a communication pathway (Latora and Marchiori, 2001). The betweenness of a node/link is given by counting the number of shortest paths between all nodes in a network over the given node/link (Freeman, 1977; Newman, 2001). The betweenness is therefore a first approximation of the traffic load in networks (Freeman, 1977) and has been investigated in various
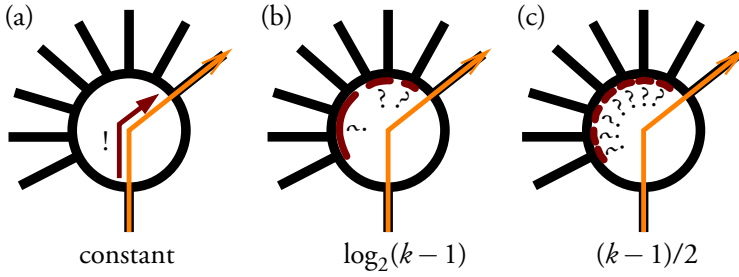
Figure 4.1: *Average information cost in bits to choose correct exit link at a node with k links depending on the possibilities at the node. (a) corresponds to a default choice where no questions are necessary. The links are ordered in some way in (b) and it is possible to successively exclude half of the links with every yes-no question. The links are not ordered in (c) and must be excluded one by one.*

contexts. For example, to model congestion in traffic networks (Holme and Kim, 2002; Holme, 2002), to model information flow in social networks (Wu *et al.*, 2004), and to establish the connection between communication and the network structure (Holme, 2003; Tyler *et al.*, 2003). However, none of them quantifies the amount of information it takes to use one of the shortest paths.

In the remaining part of this chapter I will present our approach to measure this information cost. In Paper IV we introduced three communication measures. Paper V presents an information perspective of navigation in cities, and in Paper VI we quantified the information horizon in networks. Paper VII investigates the interplay between navigation with limited information and the network structure, and in Paper VIII we summarized all findings.

Summary of paper IV

Unspecific signaling, like virus spreading or broadcasting of advertisements has been thoroughly studied (Balthrop *et al.*, 2004; Newman *et al.*, 2002; Newman, 2003b; Holme, 2004a). The goal of Paper IV was instead to quantify the information cost of *specific* signaling in complex networks. We developed a context-independent framework to answer general questions like: how much information is enough to navigate between two streets in a city, or for an e-mail to reach its recipient on the Internet? The problem goes back to Stanley Milgram's celebrated small-world experiment (Milgram, 1967), described in Chapter 3.

Every recipient, not being the target, must choose one of its friends as a new recipient of the letter. Consider a recipient i with k_i friends, and assume that its friends can be ordered in some way to make the exclusion of friends to receive the letter efficient. The procedure can for example follow the pattern: "Could it be one of my female friends?", "Could it be someone in my home town?", and so on (see Figure 4.1). Every asked yes-no question will in this way optimally reduce the possible outcomes by a factor 2. Accordingly, to find a specific friend to send the letter to one must ask $\log_2 k_i$ yes-no questions, or equivalently, $\log_2 k_i$ bits of information are necessary ($\log_2(k_i - 1)$ if the possibility to send back the letter is excluded as in Figure 4.1). The total number of bits $S(s \to t)$ used for such a letter from a source s to a target t along a shortest path can be calculated by summing up the contributions from every recipient along the path.

To generalize for *degenerate shortest paths* (two or more different paths have the same length), we defined the *search information* $S(s \to t)$ to be

$$S(s \to t) = -\log_2 \left(\sum_{\text{path}(st)} P\{p(s,t)\} \right), \qquad (4.1)$$

where the sum runs over all degenerate paths path(st) that connect s with t. $P\{p(s,t)\} = \frac{1}{k_s} \prod_i \frac{1}{k_i - 1}$ is the probability to follow such a path by chance ($k_i - 1$ to take into account the knowledge about the node already visited).
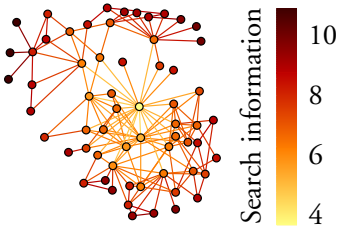

Figure 4.2: *The terrorist network from Chapter 2 colored according to the information necessary to locate a terrorist ($\mathcal{H}$ in bits).*

The search information can be averaged over all targets t for a given source s in a network to an ability of s to *seek* (or access) $\mathcal{A}_s = \frac{1}{n} \sum_t S(s \to t)$. Conversely, to average over all sources for a target corresponds to the ability of t to *hide* $\mathcal{H}_t = \frac{1}{n} \sum_s S(s \to t)$. The former is the average number of questions necessary to reach a random target t from the given source s and the latter the average number of questions necessary to locate the given target t from a random source s in the network (see Figure 4.2). By averaging over both s and t, $S = \frac{1}{n^2} \sum_{s,t} S(s \to t)$ becomes a global measure of searchability or navigability in the network.

We also investigated two measures related to predictability of global network communication from the node level. The idea was to capture the complexity of traffic routing in the nodes. For example, if a node has fifty links but only
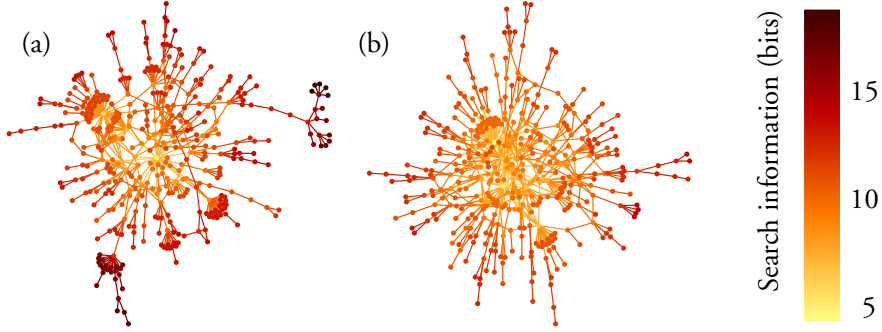
Figure 4.3: *Search information in a biological network and its randomized counter-part. The real core of the protein-interaction network of yeast (Ito* et al.*, 2001) (a) and its randomized counterpart (b) labeled by the $\mathcal{H}$, the average number of question needed to locate the colored node. Together they demonstrate the that the real network is designed toward a robust topology against noise at the expense of global accessibility.*

receives signals from one of the links and forwards them to another specific link, the complexity of the processing is low. We used the betweenness as a proxy for traffic in the networks. The betweenness, as stated before, is defined by letting every node send one signal to every other node and tracing this signal along the shortest path. This gives a density of signals on every link and we calculated the entropy of the betweenness b_{ij} around a given node i with links j and repeated this for c_{ij}, where all signals are targeted to node i,

$$R_i = -\sum_{j=1}^{k_i} b_{ij} \log_2(b_{ij}), \qquad (4.2)$$

$$T_i = -\sum_{j=1}^{k_i} c_{ij} \log_2(c_{ij}). \qquad (4.3)$$

We used these measures on a few real-world communication networks, the Internet, two social netwoks, and two biological networks. By using their random counterparts as reference points, we found that in particular the social networks proved to have predictable communication pathways (low T and R), but at the same they showed a high S meaning that they are on average inefficient in transmitting information. Conversely, the communication around most nodes

in the Internet is homogeneously distributed. The two biological networks did not have as clear deviations from their random counterparts as the other networks, possibly meaning that specific communication in biological networks is not a general process. Instead, biological networks are designed to avoid exposure to noise by their modular structure and separation of hubs (see Figure 4.3 and Paper I).

4.3   An application: Navigation in cities

When was the last time you were lost in a city? Was the event a coincidence or could it possibly be that some cities are easier to get lost in than others? How should a city be organized to make navigation easy? The aim of Paper V was to answer these questions.

However, first I would like to say a few words about cities and networks. It may seem most natural to convert a city map into a network by representing intersections with nodes and the connecting streets with links. Nevertheless, this *primal representation* has been overshadowed by one particular formulation of the opposing *dual representation*—the axial map (Hillier and Hanson, 1984). The axial map, developed in the notion of *Urban Design*, is constructed by first piecewise straighten the streets into line of sights, *axial lines*. The axial lines are thereafter turned into nodes, and each intersection between any pair of axial lines is turned into a link.

Especially the geographical centrality aspects have been extensively investigated, both of the primal and the dual mapping as well as a mapping where geographical distances were kept, in a series of papers Crucitti *et al.* (2006a); Scellato *et al.* (2006); Crucitti *et al.* (2006b). The main finding was cities can be divided into two groups with different characteristics: historical (self-organized) cities and planned cities. In particular, the organization of the streets in the historical cities obeys scale invariance. The scale invariance was also found in some investigated road networks, which was explained by the fractal placement of the roads (Kalapala *et al.*, 2006).

Summary of paper V

The approach we have used to create an *information city network* is neither of the two presented, but the method by Jiang and Claramunt (2004). We mapped street names to nodes and connected two nodes by a link if it was possible to go directly between the corresponding streets. In this way we created an informa-
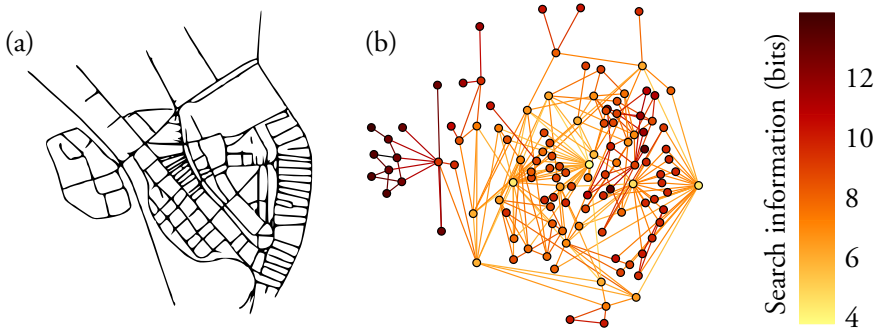
Figure 4.4: *Gamla Stan in Stockholm (a) mapped to an information city network (b). Nodes are roughly positioned at the geographical position of the corresponding street and color coded according to the typical amount of information needed to locate them (H from Paper IV on page 34).*

tion representation of a city as in Figure 4.4. We used S, $\mathcal{H}$ and $\mathcal{A}$ from Paper V to estimate the navigability in some Swedish cities (Stockholm, Gothenburg, Malmö and Umeå) together with Manhattan as a representation of a modern planned city. This approach makes sense if you are a newcomer in the city, but the measure is also equivalent to the amount of information gained by knowing a city. Another interesting aspect is that $S(s \rightarrow t)$ estimates the mental feeling of asymmetry when traveling a way back and forth. The first way always feels longer, possibly due to the increased level of consciousness associated to the navigation.

We measured basic quantities like degree distribution and clustering to understand why some cities are more difficult to navigate in than others (also after taking size effects into account). We found an excess of triangles that limits the navigability because triangles give more choices at every street (higher information cost), but do not contribute to degenerate shortest paths leading to the target. From Paper IV we know that degenerate shortest paths facilitate navigation, and Manhattan with its excess of squares is therefore a good example of how a favorable city looks like from the perspective of navigation.

Relatively high values of S, $\mathcal{H}$ and $\mathcal{A}$ compared with the random counterpart indicate that cities are not generally designed for easy navigation, but rather to avoid exposure—to hide. We do not want highways outside our windows, even if that would make navigation easier!

4.4   The information horizon

The larger the network the better the statistics. This simple fact leads us inexorably to look for large networks. However, let us pretend that we can construct a network of all people in the world and then ask: "What is my relation to a bookseller in Kabul or the sultan of Brunei?" What meaning does the answer have from a system point of view? The scope of Paper VI was to define a horizon at the edge of where direct and efficient communication turn into limited and indirect communication.

Summary of paper VI

In Papers IV and V we found that the search information S is higher in most real networks than in their respective random counterparts. However, this is not true for navigation at all shortest path lengths l. In Paper VI we investigated $S(l)$ and found that it is often easier to navigate on short distances in real-world networks compared with their random counterparts. This horizon that separates easy navigation from relatively more difficult navigation was found to be at a distance $l \leq 3$ away from the nodes. We interpreted this as a design toward efficient local communication, at the expense of a need for more intelligent methods to facilitate searchability on long distance. This should also be interpreted as an organization to avoid exposure to nonspecific communication, that is noise.

We investigated a few different methods to improve navigability on long distance, all based on the same idea to use traffic as a way to ask smart questions while navigating to a distant target. For example, a good idea when one is lost in a city is to follow the main flow and benefit from the fact that all participants in the traffic flow also have a target. In this way we defined a traffic-weighted search information (see Figure 4.5)

$$S_w(s \rightarrow t) = -\log_2 \left(\sum_{\{p(s,t)\}} b_{s,j=1} \prod_{j \in p(s,t)} b'_{j,j+1} \right), \tag{4.4}$$

with the betweenness b as an estimate for traffic (b' eliminates the traffic from the already visited node). In this way the number of questions is reduced. In fact, we found that this method is almost optimal when navigation across the full investigated networks. However, one has to bear in mind that we have neglected the cost associated to estimating the traffic flow.

Back to Stanley Milgram's six degrees of separation experiment, one can speculate in how a recipient did to choose one of its friends as the best candidate to

deliver the letter as fast as possible. More generally, how do we do when we have to locate a remote node in a representation of some aspect of the world we live in.

The nontrivial result of the experiment is not that the distance between two persons is only six, since the dimension in social networks is high (Kochen, 1989), but the fact that the short paths were found in the experiment. We speculated that an estimate of the distance to the target must facilitate the search. This information was available, since the letters indicated in which city the target person was living in. To estimate the reduction in the number of necessary questions this information gives, we restricted the betweenness in Equation (4.4) to only be traffic between nodes separated by at most the remaining distance to the target. For example, we

Figure 4.5: *The optimal search strategy on the information city network of Gamla Stan, Stockholm.*

found that while navigating in a city, it is a good choice to first use the cars as an estimate for traffic and while approaching the target go on to use bicycles and later pedestrians as estimates (see Figure 4.5).
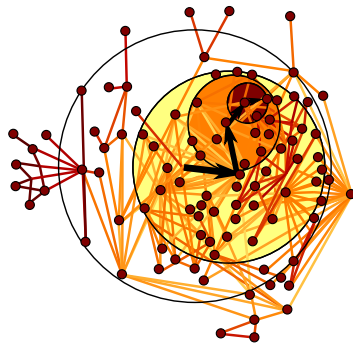
4.5 Limited information

In Papers IV-VI we assumed that the necessary information associated with the navigation is available. Nevertheless, to stand on a street in an unknown city and long for the hotel bed without anyone to ask for the way, is more characteristic of the real world than well-prepared guides in every corner. As the real world is best characterized from a limited information perspective, we generalized the measure S to a context with limited information in Paper VII.

Summary of paper VII

If the available information is not enough to walk a shortest path from a source s to a target t, it is natural to remove the restriction that an already visited node can not be visited again. Then $k - 1 \rightarrow k$ in the denominator of Equation (4.1). Further, as the shortest path no longer is a good estimate of the actually
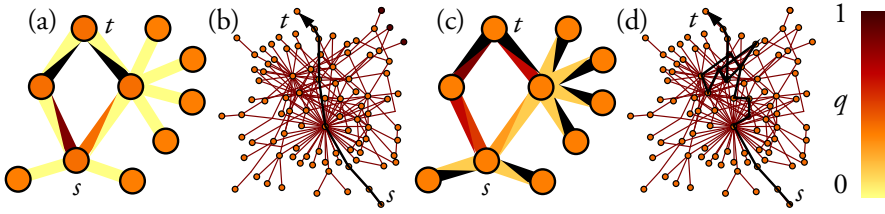
Figure 4.6: *Search with limited information. At each node i along a path from source s toward target t, the information cost is unlimited in (a), but limited to ι = 0.2 bits at every node in (c). As a consequence the total information cost is higher in (a) than in (c), on average 2.6 bits in (a) compared to 0.7 bits in (c), but the walked path is on average 2.5 steps shorter in (a) according to simulations. The color of the links represent the probability q to leave a node along a link on the path toward t. (b) shows a typical navigation with unlimited access to information and (d) with limited (ι = 1.0 bits) information.*

walked path, which can now be very different from time to time if the available information is low, the search information must be defined locally. This is simple if there are no degenerate paths. The cost ι_{it} at node i on the path to the target node t is simply $\iota_{it} = \log_2(k_i)$. Let the total information now be I, then $I(s \rightarrow t) = \sum_{i \in \mathrm{path}(st)} \iota_{it}$.

This means that the available information is unlimited at every node, $\iota = \infty$, like in Papers IV-VI ($I_\infty(s \rightarrow t) \approx S(s \rightarrow t)$[1]). This was generalized for degenerate paths by choosing probabilities q for every exit link at a node to minimize the total information needed to go from s to t (see Figure 4.6(a) and (b)). On the other hand, if the upper limit of available information at a node is zero ($\iota = 0$), the probability to take a specific link is as high as any other link and the walk is random but without any information cost ($I_0(s \rightarrow t) = 0$). To be able to study intermediate ι we smeared out the probabilities to take a given link from a node i to satisfy $\iota_i = \iota$ as in Figure 4.6(c) and (d).

In Figure 4.6(c) the information at any node is limited to $\iota = 0.2$ bits. The probabilities to take a given link have shifted from Figure 4.6(a) so that there is now a chance to make a mistake and follow a link that is not on the shortest path to t. This scenario gives rise to an interplay between navigation and limited information, and the scope of Paper VII was to quantify this interplay.[2]

---

[1]The difference comes from whether k or $k-1$ options are considered at a node.
[2]A simulation is available on http://cmol.nbi.dk/javaapp.php

We compared some real-world networks with their random counterarts to see if the real-world networks have a topology that decreases the effects of the limited information. Again, and even more pronounced, we found that many real-world networks are designed to favor communication on short distance at the expense of communication on long distance. By studying model networks with different topological features, we found that it is especially the modular structure of the networks that gives rise to this horizon. In other words, some more or less well defined parts are highly integrated and separated from other parts of the network. The effect was even more striking if the different modules were interconnected by a backbone of highly connected nodes. This feature was for example observed for the Internet where modules naturally occur within countries and highly connected autonomous systems connect the countries.

Summary of paper VIII

In Paper VIII we summarized the findings covered in this chapter, but also introduced a generalized network hierarchy measure. The search information $S(s \rightarrow t)$ makes it possible to in detail measure the relative importance of nodes in a given network. In fact, measuring the *visibility* of a node t in terms of how well hidden the node is from the rest of the network, we showed how networks can be ranked in terms of a generalized hierarchy measure. The measure captures both the hierarchy in the usual terms of trees like in a military organization, and at the same time also the intrinsic topological hierarchical nature of scale-free networks (Trusina *et al.*, 2004). This generalized hierarchy measure defines scale-free networks with degree distributions with exponents close to $\gamma = 2$ to be hierarchical, whereas narrower distributions will be antihierarchical unless they are deliberately organized in a treelike structure.

4.6   Beyond the information horizon

The different ways networks can be organized were in this chapter reflected on their ability or inability to transmit specific messages across the networks. The surprising result that almost all real-world networks on average are more difficult to navigate than their random counterparts was explained by the information horizon: the topology of the networks favors communication at short distance, but disfavors navigation at larger distances beyond the horizon. The horizon, in turn, was explained to be in part an effect of the modular structure of the

networks, and the way the modules are interconnected. However, very little was said about why networks are modular.

The emergence of modules in social networks will be discussed in the next chapter. However, it is easy to understand the modular structure in cities. Here the modules to a high degree consist of suburbs that have merged together by growth. Every module is in itself a working unit like the organelles in a cell or the organs in our body. This takes us to the modularity in the biological networks, and with that to the inevitable connection to evolution. The biological networks have evolved by small changes, mutations, that most often were deleterious, but sometimes resulted in more fit organisms. Obviously, a successful organism is an organism with an intrinsic structure such that the small changes inferred by mutations result in small functional changes of the organism. One way of achieving this is to separate functions and make the network modular (Smith and Szathmáry, 1995). The modular network that we see is therefore nothing but a network of networks.

Chapter 5

Communication and the information horizon

A s humans most of us chat about anything or about anybody at any time, but not with anyone. Instead we have a set of acquaintances, with whom we communicate on a regular basis. All the local interactions form together a communication network over which information can be transmitted globally.

This everyday chatting is not just chatting for the sake of chatting. Instead we often strive for new information of the kind that can not be acquired by reading just any mainstream newspaper or watching public TV (or when you get the information this way it is already too late).[1] In this way, everyone's well-being is dependent on the actions of one's own as well as the actions taken by friends (Coleman, 1966; Granovetter, 1994), let it be the ability to get a job or to have access to the latest fashion.

The requested information is not always locally accessible, but must be acquired by sequences of local communication events across the network. The following study is therefore not only a study of communication in a social context, but also a general framework to understand the interplay between the network topology and the overall ability to organize communication in a system. Our approach should therefore not be confused with neither the informational theoretical approach of Shannon to quantify the amount of information transferred in a signal between a sender and a receiver (Shannon and Weaver, 1949), nor

---

[1]The general rule is: When you hear a taxi driver talk about subscribing for shares it is already too late to go to the stock market.

with communication in technological networks where queuing and congestion are the central issues (Guimerà *et al.*, 2002).

5.1   Modeling social communication

We know from the preceding chapter that it is in principle easy to organize global communication, given that the network structure is known. However, this knowledge about the global structure is absent on local level in most systems (Milgram, 1967; Watts *et al.*, 2002; Kleinberg, 2000a). There is no network administrator to turn to in social networks and everyone will therefore try to make a rough picture of the network from their position. The communication in networks must therefore play double roles, both as a transmitter of information and ideas (Daley and Gani, 2000; Jackson and Yariv, 2005), but also as a tool for individuals to try to build a global *perception* of the network and thereby make it possible to overcome the information horizon set by the immediate neighbors (Friedkin, 1983). When I hesitate whether I should use my slim-fit jeans or not, five minutes before the taxi comes to take me to a public ceremony, I can not call all 200 people in my phonebook. Instead I must use my locally acquired and limited global knowledge of my social network, and call the one that actually happens to have a friend who knows a stylist. I can use my perception, my rough picture of the network, because I have recently talked to this particular friend about fashion, which led us to talk about the stylist (see Figure 5.1).

  This example highlights not only the dual roles of communication, but also its limited local nature. Communication is a torch that is stuck to the nodes in the network and can not shed light on the network from above. We know of course more about people in our neighborhood in the network than distant people, and more about people who communicate a lot than reserved persons (Friedkin, 1983; Kumbasar *et al.*, 1994; Bondonio, 1998).

  The challenge is to model this pattern of communication taking place on the network, which forms local individual perceptions that in turn will affect the pattern of communication. Communication habits as well as the network topology will therefore play fundamental roles in this game of access to information through communication.

  Our approach was to use agents as players in this social game. The agents' limited perception of the network was modeled by giving every agent a memory (see the memory bubble of agent **A** in Figure 5.1). Every agent therefore has, for every other agent in the system, a pointer to an immediate friend that it believes
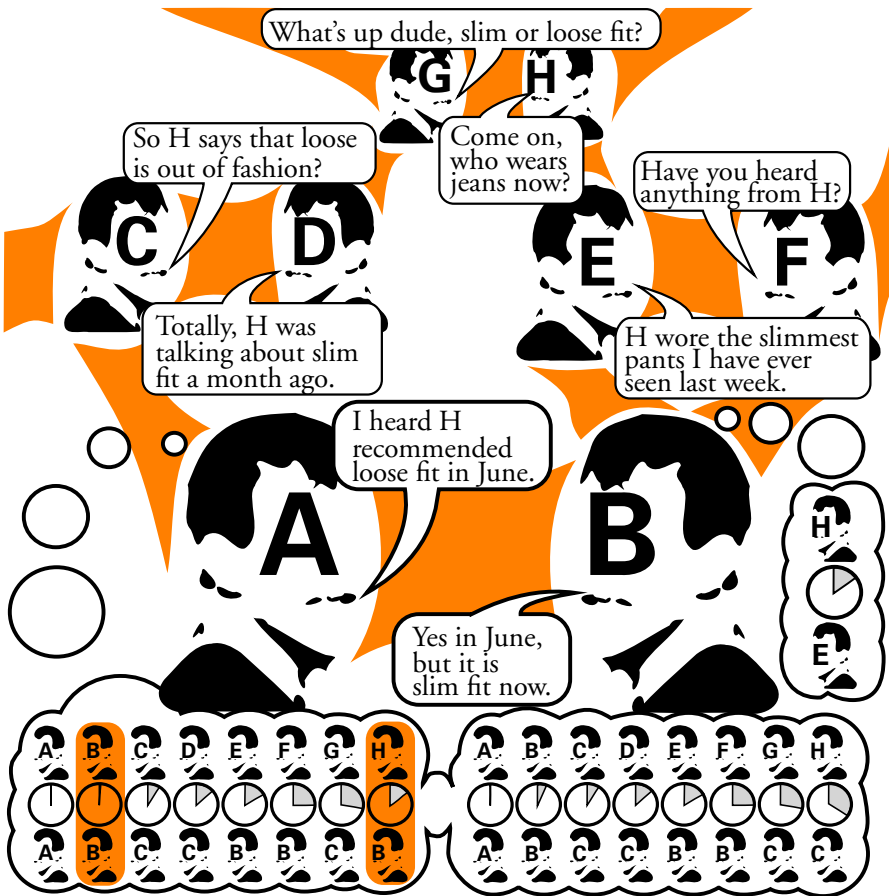
Figure 5.1: *Snapshot of a communication model when everyone talks about agent H. By local communication we can not only transmit information over distance, but we can also create a rough picture of our social network. In our agent-based models, this perception is a memory for every agent that consists of pointers to other agents in the network. Here agent A has different perception before (right memory bubble) and after (left memory bubble) the communication event with agent B. For every agent (top row in memory bubbles), the agent has a pointer in the form of a nearest neighbor (bottom row in memory bubbles), assuming that it has better access to the other agent. Additionally the agents have a metric that estimates the quality of this information (middle row in memory bubbles, here a clock representing the age of the information). In this example, as A has talked to B about H, A updates its perception about B and H (see difference between right and left memory bubble of A).*

can provide the best access to the other agent. Look at Figure 5.1 and agent **A**. Agent **A** has two nearest neighbors in the network consisting of eight agents, the agent itself included, and will therefore have seven pointers distributed between its two friends **B** and **C**, and one pointing to itself (because the agent has better information about itself than any other agent).

   We let the individual perceptions become updated in local communication events modeled by letting two neighbors talk about a third person. For example Bergmann (1993) argues that a substantial part of our everyday conversation takes this form, and we believe that it is a good approximation. When the two agents communicate, they compare the quality of their information about the third agent with each other and can therefore estimate who has best access to the information. By this we do not mean that every conversation is about competing in dirty details about other people, but that it is an inherent component of everyday conversation to compare the validity of the information. A comparison that often results in one part proving to have better access to the third person, and thereby defining a direction to that person. This pointer is the fundamental unit in our modeling of a local perception of the network.

5.2   Modeling social mobility

By highlighting that our information horizon is set by each individual's social contacts in a society, which in turn are parts of the global network of human communication, we capture the coupling between access to information and the position in the network. We also capture the important role the topology plays for the overall accessibility, as well as the coupling between communication habits and the ability to get access to new information (see Paper X). However, we lack one important component. Communication is not the only way to get access to information across a network. As a person is limited to interact with her neighbors, the possible success therefore not only depends on her action, but also on the actions taken by her neighbors, her neighbors' neighbors, and so on. Why not take advantage of this coupling between success and the position in the network and strive toward a good position?

   This social mobility can be demonstrated by the action I take the next time I face the dilemma of slim or loose-fit jeans. I will contact my friend's friend directly to get even more up to date information (see Figure 5.2). That I will not just contact anyone randomly highlights the limited social mobility we have. Instead the new friendship has been established based on my locally acquired in-
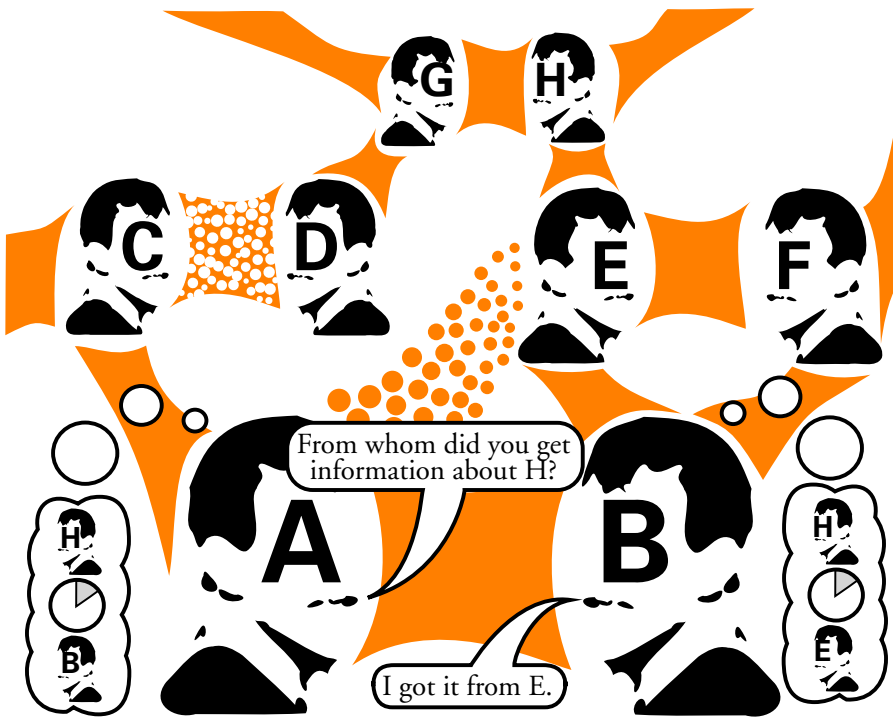
Figure 5.2: *A model of social mobility. By successive rewiring attempts the agents try to optimize their positions and minimize the distances to other agents. They do this based on the locally acquired information they have got through communication with their neighbors (see Figure 5.1). In this case agent **A** wants to be closer to agent **H**, and use its perception as well as the perception of its friend **B** to achieve this.*

formation about the network, which in turn is based on communication over the same network. The change in the network will change my, as well as others, perception of the network, because the network over which the communication takes place is different. Different perceptions will lead to different actions, and all together result in an interesting self-organized feedback game between communication habits and the topology. In simple terms (Newman *et al.*, 2006): "The ties people make affect the form of the network, and the form of the network affects the ties people make."

The aim with the three papers that will be presented in this chapter has been

to make realistic, yet very simple, models of this network game that we play on an everyday basis (Carpenter *et al.*, 2003; Friedkin, 1982).[2] By doing this, we present a model framework that couples *dynamics on* the network with the *formation of* the network.

Summary of paper IX

Paper IX was our first attempt to explore the local dynamic origin of global network organization by modeling response to information transfer in a simplified social system. As described above, the scenario was a set of players, who each tried to be as close as possible to everybody else. The players adjusted their social connections to achieve this goal. They were guided by a limited knowledge about the other players' positions in the network (Figure 5.2), which was obtained by local communication (Figure 5.1).

Before I presented the communication step and the rewiring step (the addition of a link to a friend's friend) separately, but in this first model we merged the two steps into one. We also used distance as a quality identifier instead of time in the following two papers.

Nevertheless, we were able to show that when the local communication is weak, the system disorganized into a highly dynamic and chaotic network where no single player dominated the system. In network language, the degree distribution was random and narrow. On the other hand, when local communication was strong, the system organized into a coherent structure dominated by a central hub that remained indefinitely frozen. In between the weak and strong local communication, there was a critical transition in the dynamics of the network where no hubs took over for ever, and where at the same time the network had players with all types of connectivities. The network was scale-free and furthermore hierarchical (Trusina *et al.*, 2004), in a way that resembled the Internet and some social and biological networks. In this perspective, we interpret scale-free networks not as an extension of random networks with a broad degree distribution forced on, but as a collapse of a centralized system with large hubs. In the presence of limited information the system can not sustain its centralized structure, but disperse into a distributed system with a power-law degree distribution half way to total randomness

The social communication model opened for investigation of the interplay between individual behavior and global organization, as well as for exploring

---

[2]Java simulations of all models are available on http://cmol.nbi.dk/javaapp.php

possible success strategies for individuals. For example, we found that scale free-networks may be associated to a dynamics on the edge of chaos. In fact, the system can self-organize at the transition between the frozen and the disordered state by a simple feedback mechanism associated to just the two most connected players in the system. Another interesting feature is that the success of individuals appears not so much to depend on their correctness, but rather on their ability to communicate actively. A talkative player's boasting is a self prophecy in the sense that it will help it to become a major player in the system. In other words, name-dropping pays off, also in a simulated society!

Summary of paper X

In contrast to Paper IX where we analyzed both the communication and the network dynamics, we focused exclusively on the communication in Paper X. The reason was that we wanted to make a clear cut presentation of what we believe is a very interesting model. Moreover, in this model we used time instead of distance as a proxy for the quality of the information. This is exactly what is shown in Figure 5.1, and the rule is very simple: Adopt information from your neighbor if it provides you with newer information than you already have. For example, if **A** talks to **B** about **H**, **A** finds that **B** has more recent information, and understands that **B** probably is closer to **H**. **A** will thereafter associate **B** as the acquaintance to turn to, to get new information about **H**. That is, **A** changes the information associated to **H** by copying the age from **B** and by pointing to **B** (see the two memory bubbles before and after the communication event in Figure 5.1). By repeated small talks like this the agents will create a perception of the whole without talking with everybody, but only with connected acquaintances in the network.

In this minimalistic social communication model, we succeeded to address the question of trade off between robustness and efficiency of communication in a realistic framework. We used scale free and ER-networks as model networks with and without hubs and modeled noise by randomly rewiring links at different rates. We found that messages are most effectively forwarded in the presence of hubs in a noise-free environment, but showed that scale-free networks are less robust signal transmitters than networks without hubs in a noisy environment. Hubs are sometimes efficient, but they tend to accumulate mistakes, and are not necessary to provide short paths. On the other hand, networks without hubs provide many alternative short paths. Maybe most importantly, the success of scale-free networks is based on the activity of the hubs. The activity of the nodes

in the scale-free network must be proportional to their number of links to make use of the hub structure.

Finally, the framework opened up for schematic modeling of self-assembly of information in a variety of systems, including social systems. This is the topic of the next paper.

Summary of paper XI

In the model presented in Paper X, we showed that it is possible to build a reliable perception of the whole, and thereby overcome the information horizon set by the neighbors, through repeated small talks. We simply let agents memorize the acquaintances that provided the newest information about other agents together with the age of this information. In Paper XI we took that model one step further and gave the agents a social mobility like in Paper IX, but here presented ins a simpler model. With the social mobility, the agents can get new acquaintances to meet different interests, and we have a two-step model.

- *Communication:* Select a random link and let the two agents that it connects communicate about a random third agent. The two agents also update their information about each other.

- *Rewiring:* Select a random agent and let it use the local information to ask an acquaintance about whom to establish a link to, to shorten its distance to a random agent. Subsequently a random agent loses one of its links.

For example, agent **A** in Figure 5.2 would like to be better positioned relative to agent **H** (the stylist). **A** therefore asks **B**, the acquaintance that provided **A** with the newest information about **H**, from whom **B** got the information. **B** answers **E**, and **A** establishes a contact with **E**. At the same time a random agent loses a random link to keep the number of links constant (the link between **C** and **D** in Figure 5.2). In this way the social mobility of the agents is limited to steps to an acquaintance's acquaintance.

This approach, with two independent events (communication and rewiring) allowed us to study in detail the self-organized feedback between communication and topology in social networks. We simply varied the number of communication events relative to the number of rewiring events. Again, as in Paper IX, we observed a narrow distribution of links when the communication was low and a system with a broad distribution of links when the communication was high. However, this time we got better insight in the interplay because we also quan-

tified the quality of the agents' perception at different communication rates. We did this by trying to send messages, guided by the agents' pointers, between all pairs of agents. Naturally, we found that the perception of the agents and the communication backbone differs substantially at low communication rates. The result was that most messages were lost. However, at higher communication rates almost all messages reached their targets, the rewiring became meaningful, and the network topology no longer random.

The model describes a social game where the aim is to be central, and a winner is an agent with many connections that provide short and reliable communication to other agents. This makes it tempting to investigate whether there are some particular strategies with which agents can improve their standing in the network.

We tested a few strategies, and found for example, that the more an agent chats with its surroundings, the better perform both the agent itself and the surroundings. However, extensive chatting is expensive and one might ask if there is a cheaper way around. A successful agent is an agent that provides new and updated information about other agents, as this encourages other agents to connect to it, which in turn gives the agent even shorter information paths to other agents. Why not exploit the system and lie about the age of the information about other agents? At the public party, wearing my slim fit jeans, I would probably not refer to the real age of the information from the stylist. This information could in fact be quite old because it came through two intermediate friends. Instead I could say, "Yes you know, many stylists talk about slim fit now," and hope that people would see me as a forerunner, a dandy, and thereby someone with fast access to information. We called this lying, and were able to show that only a single liar in a system with nonliars benefit from the strategy. Lying is so destructive that a few liars are enough to break down the reliability of the communication and nobody is in reality a winner.[3]

5.3   Beyond the communication horizon

Our scheme of investigating self-organization of networks under various degrees of limited information opens for a family of socioeconomic models, where gain and loss are quantified through redistribution of links in a network. The models will in their most simple form operate with globally homogeneous strategies,

---

[3]See illustrative Java simulation at http://cmol.nbi.dk/models/inforew/inforew.html

but will in more realistic scenarios include agents with an assortment of heterogeneous strategies of communication and link redistribution. I will here refer to some unpublished results to substantiate this richness of the model framework.

We saw for example in Chapter 4 that the information horizon was in part a result of the modularity of the networks. It is therefore tempting to speculate in the origin of this modularity in this framework of self-organization between communication and topology.

A simplification that we made in Papers X and XI was to distribute an agent's interest evenly among the other agents. This is of course not realistic, and we have therefore tested to modify the communication rule and let the probability to talk about someone be proportional to how often it already has talked about this agent. This is a very natural second order approximation, and the agents naturally got better perception about agents close by and worse perception about distant agents in accordance with the information horizon.

However, people also have different interests, which is reflected in the communication pattern. To further study the self-organization we divided the agents into different groups of interest. Within each group the agents keep track of every other agent, but they let the other groups be represented by a single slot in the memory. This coarse graining of each individual agent's perception of the network resulted in the emergence of modules in simulations of the system. This, more than anything else, shows the strong coupling between information, communication, and the topology in complex networks.

The natural next step to take this work further would be to add trust, reputation, and the concept of asymmetric information to the game (Akerlof, 1970). This mix of socioeconomic models and self-organized information network models would result in, I believe, a very interesting field of research.

Chapter 6

Summary

Communication is a fundamental element in maintaining the overall cooperation between distant parts of a complex system. Because signals and traffic follow specific links between sender and target, network modeling is a powerful tool to illustrate and understand the paths of and the conditions for communication.

In this thesis, we have shown how the network topology affects the communication ability in various networks, from information-city networks to biological and social networks. For example, we have confirmed the assumption that Manhattan, and presumably most modern planned cities, are simpler to navigate in than older cities with a long history of construction, by quantifying the information associated with navigation. We have also demonstrated, that many real-world networks favor communication on short distance at the expense of communication on long distance. This organization is a way to avoid exposure to nonspecific communication—noise. We call the boundary between the favored and unfavored communication, typically two-three steps away in the network, the information horizon.

The underlying topological features of this horizon were found to be a modular and often also a hierarchical organization with highly connected nodes deliberately positioned between other nodes. The information horizon is therefore not only associated with broad degree distributions, but also with well known organizational features of urban, social, and biological systems.

Because efficient information transfer is a fundamental task of many networks in nature, we have considered the dynamics of networks as an optimization process. Every action or process between the nodes is more or less complicated and time consuming. By merging nodes into bigger units these costs can

be reduced, and thereby increase the overall functionality in the network. This is the core of the merging-and-creation models presented in various contexts in the thesis, all giving rise to scale-free degree distributions.

We have also illustrated how the perception of the network, from all nodes point of view (and achieved by local communication), affects the way information is transferred between distant parts of the network. The approach was to let agents, in a network model, chat with nearest neighbors about distant agents in the network, and thereby self-organize communication pathways. We demonstrated that only a few rules were sufficient to locally create pictures of the network, and that the pictures were adaptable to changes in the network. In a society the information horizon is set by each individual's social contacts, which in turn is a part of the global network of human communication. By introducing this social game we were able to investigate different ways to expand the perception of the system and push the information horizon outward.

We also explored the local mechanism behind network organization by modeling the response to communication in simplified social systems. We showed that low communication levels result in random Erdős-Rényi networks, whereas higher communication levels, with more reliable communication pathways, lead to structured networks with broad degree distributions. Moreover, communication, even when unequally distributed, benefits every agent in the network. However, by studying agents giving false information, we found that not only the position in the network, but also the validity of the transferred information, influences if nodes are favored or not.

It is now time to reveal what Hamlet, holding a network instead of a skull in his hand, would have said:

—To connect, or not to connect: that is the question!

References

Akerlof, G. (1970), *The market for lemons: Quality uncertainty and the market mechanism*, Quart. J. Econ. **84**, pp. 485–500.

Alava, M. J. and Dorogovtsev, S. N. (2005), *Complex networks created by aggregation*, Phys. Rev. E **71**, art. no. 036107.

Albert, R., J., H., and Barabási, A.-L. (1999), *Diameter of the World-Wide Web*, Nature **401**, pp. 130–131.

Albert, R., Jeong, H., and Barabási, A.-L. (2000), *Error and attack tolerance of complex networks*, Nature **406**, pp. 378–382.

Amaral, L. and Ottino, J. (2004), *Complex networks: Augmenting the framework for the study of complex systems*, Eur. Phys. J. B **38**, pp. 147–162.

Amaral, L. A. N., Scala, A., Barthélémy, M., and Stanley, H. E. (2000), *Classes of small-world networks*, Proc. Natl. Acad. Sci. USA **97**, pp. 11149–11152.

Ball, F., Mollison, D., and Scala-Tomba, G. (1997), *Epidemics with two levels of mixing*, Ann. Appl. Probab. **7**, art. no. 1, pp. 46–89.

Balthrop, J., amd M. E. J. Newman, S. F., and Williamson, M. M. (2004), *Technological networks and the spread of computer viruses*, Science **304**, pp. 527–529.

Barabási, A.-L. (2001), *The physics of the Web*, PhysicsWorld .

Barabási, A.-L. and Albert, R. (1999), *Emergance of scaling in random networks*, Science **286**, pp. 509–512.

Barabási, A.-L. and Albert, R. (2002), *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74**, pp. 47–98.

Barabási, A.-L., Albert, R., and Jeong, H. (1999), *Mean-feld theory for scale-free random networks*, Physica A **272**, pp. 173–187.

Barabási, A.-L., Albert, R., Jeong, H., and Bianconi, G. (2000), *Power-law distribution of the world wide web*, Science **287**, p. 2115.

Barrat, A. and Weigt, M. (2000), *On the properties of small-world network models*, Eur. Phys. J. B **13**, pp. 547–560.

Bergmann, J. R., *Discreet Indiscretions: The Social Organization of Gossip* (Aldine De Gruyter, New York, 1993).

Boccaletti, S., V. Latora and, Y. M., Chavez, M., and Hwang, D.-U. (2006), *Complex networks: Structure and dynamics*, Phys. Rep. **424**, pp. 175–308.

Boguñá, M. and Pastor-Satorras, R. (2002), *Epidemic spreading in correlated complex networks*, Phys. Rev. E **66**, art. no. 047104.

Bondonio, D. (1998), *Predictors of Accuracy in Perceiving Informal Social Networks*, Soc. Networks , pp. 301–330.

Bornholdt, S. and Schuster, H. G., eds., *Handbook of Graphs and Networks: From the Genome to the Internet* (Wiley-VCH, Germany, 2003).

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000), *Graph structure in the web*, Comput. Netw. **33**, pp. 309–320.

Caldarelli, G., Capocci, A., de los Rios, P., and Muñoz, M. A. (2002), *Scale-free networks from varying vertex intrinsic fitness*, Phys. Rev. Lett. **89**, art. no. 258702.

Carpenter, D., Esterling, K., and Lazer, D. (2003), *The strength of strong ties: A model of contact-making in policy networks with evidence from U.S. Health Politics*, Ration. Soc. **15**, pp. 411–440.

Carroll, L., *Sylvie and Bruno Concluded* (Macmillan and Co, London, 1893).

Cohen-Cory, S. (2002), *The developing synapse: Construction and modulation of synaptic structures and circuits*, Science **298**, pp. 770–776.

Coleman, J., *Medical Innovation: A Diffusion Study* (Bobbs-Merrill, New York, 1966).

Crucitti, P., Latora, V., and Porta, S. (2006a), *Centrality in networks of urban streets*, CHAOS **16**, art. no. 015113.

Crucitti, P., Latora, V., and Porta, S. (2006b), *Centrality measures in spatial networks of urban streets*, Phys. Rev. E **73**, art. no. 036125.

Daley, D. J. and Gani, J., *Epidemic Modelling* (Cambridge University Press, Cambridge, UK, 2000).

Dennis, B. R. (1985), *Solar hard X-ray bursts*, Solar Phys. **100**, pp. 465–490.

Dorogovtsev, S. and Mendes, J. (2002), *Evolution of networks*, Adv. Phys. **51**, pp. 1079–1187.

Erdős, P. and Rényi, A. (1959), *On random graphs I*, Publ. Math. Debrecen **6**, pp. 290–297.

Faloutsos, M., Faloutsos, P., and Faloutsos, C., *On power-law relationships of the Internet topology*, in *SIGCOMM* (1999), pp. 251–262.

Fell, D. A. and Wagner, A. (2000), *The small world of metabolism*, Nat. Biotechnol. **18**, pp. 1121–1122.

Field, G. B. and Saslow, W. C. (1965), *A statistical model of the formation of stars and interstellar clouds*, Astrophys. J. **142**, p. 568.

Freeman, L. C. (1977), *A set of measures of centrality based on betweenness*, Sociometry **40**, pp. 35–41.

Friedkin, N. E. (1982), *Information flow through strong and weak ties in intraorganizational social networks*, Soc. Networks **3**, pp. 273–285.

Friedkin, N. E. (1983), *Horizons of observability and limits of informal control in organizations*, Soc. Forces **62**, pp. 54–77.

Fry, D. B., *Homo loquens: Man as a talking animal* (Cambridge: CUP, 1977).

Gingerich, O., *The Book Nobody Read: Chasing the Revolutions of Nicolaus Copernicus* (Walker Publishing Company, Inc, 2004).

Girvan, M. and Newman, M. E. J. (2002), *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. **99**, pp. 8271–8276.

Goh, K.-I., Kahng, B., and Kim, D. (2001), *Universal behavior of load distribution in scale-free networks*, Phys. Rev. Lett. **87**, art. no. 278701.

Granovetter, M., *Getting a Job: A Study of Contacts and Careers* (Northwestern University Press, Evanston, 1994).

Grönlund, A. and Holme, P. (2004), *Networking the seceder model: Group formation in social and economic systems*, Phys. Rev. E **70**, art. no. 036108.

Grönlund, A. and Holme, P. (2005), *A network-based threshold model for the spreading of fads in society and markets*, Adv. Complex Syst. **8**, pp. 261–273.

Grönlund, A., Sneppen, K., and Minnhagen, P. (2005), *Correlations in networks associated to preferential growth*, Phys. Scr. **71**, pp. 680–682.

Guimerà, R., Díaz-Guilera, A., Vega-Redondo, F., Cabrales, A., and Arenas, A. (2002), *Optimal network topologies for local search with congestion*, Phys. Rev. Lett. **89**, art. no. 248701.

Hillier, B. and Hanson, J., *The Social Logic of Space Cambridge University Press* (Cambridge University Press, Cambridge, UK, 1984).

Holme, P. (2002), *Edge overload breakdown in complex networks*, Phys. Rev. E **66**, art. no. 036119.

Holme, P. (2003), *Congestion and centrality in traffic flow on complex networks*, Adv. Complex Syst. **6**, pp. 163–176.

Holme, P. (2004a), *Efficient local strategies for vaccination and network attack*, Europhys. Lett. **68**, pp. 908–914.

Holme, P. (2004b), *Form and function of complex networks*, Ph.D. thesis, Department of Physics, Umeå university, 901 87 Umeå.

Holme, P. and Kim, B. J. (2002), *Vertex overload breakdown in complex networks*, Phys. Rev. E **65**, art. no. 066109.

Hughes, D. and Paczuski, M. (2003), *Scale-free magnetic networks: Comparing observational data with a self-organizing model of the coronal field*, Preprint astro-ph/0309230 .

Hughes, D., Paczuski, M., Dendy, R. O., Helander, P., and McClements, K. G. (2003), *Solar flares as cascades of reconnecting magnetic loops*, Phys. Rev. Lett. **90**, art. no. 131101.

Ising, E. (1925), *Beitrag zur Theorie des ferromagnetismus*, Z. Physik **31**, pp. 253–258.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001), *A comprehensive two-hybrid analysis to explore the yeast protein interactome*, Proc. Natl. Acad. Sci. U.S.A. **98**, pp. 4569–4574.

Jackson, M. O. and Yariv, L. (2005), *Diffusion on Social Networks*, Économie Publique **16**, pp. 3–16.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A.-L. (2000), *The large-scale organization of metabolic networks*, Nature **407**, pp. 651–654.

Jiang, B. and Claramunt, C. (2004), *Topological analysis of urban street networks (small-world or not)*, Environ. Plann. B. **31**, pp. 151–162.

Kalapala, V., Sanwalani, V., Clauset, A., and Moore, C. (2006), *Scale invariance in road networks*, Phys. Rev. E **73**, art. no. 026130.

Katz, L. and Powell, J. H. (1957), *Probability distributions of random variables associated with a structure of the sample space of sociometric investigations*, Ann. Math. Stat. **28**, pp. 442–448.

Kim, B. J., Trusina, A., Minnhagen, P., and Sneppen, K. (2005), *Self-organized scale-free networks from merging and regeneration*, Eur. Phys. J. B **43**, pp. 669–672.

Kleinberg, J. (2000a), *Navigation in a small world*, Nature **406**, p. 845.

Kleinberg, J., *Small-world phenomena and the dynamics of information*, in *Proc. 32nd ACM Symposium on Theory of Computing* (2000b).

Klemm, K. and Eguíluz, V. M. (2002), *Growing scale-free networks with small-world behavior*, Phys. Rev. E **65**, art. no. 057102.

Kochen, M., *The small world* (Ablex, Norwood, 1989).

Krapivsky, P. L. (1993), *Aggregation-annihilation processes with injection*, Physica A **198**, pp. 157–178.

Krapivsky, P. L. and Redner, S. (2002), *A statistical physics perspective on Web growth*, Comput. Netw. **39**, pp. 261–276.

Krebs, V. E. (2002), *Mapping networks of terrorist cells*, Connections **24**, art. no. 43–52.

Kumbasar, E., Romney, A. K., and Batchelder, W. (1994), *Systematic Biases in Social Perception*, Am. J. Sociol. **100**, pp. 477–505.

Latora, V. and Marchiori, M. (2001), *Efficient behavior of small-world networks*, Phys. Rev. Lett. **87**, art. no. 198701.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Åberg, Y. (2001), *The web of human sexual contacts*, Nature **411**, p. 907.

Lin, R. P., Schwartz, R. A., Kane, S. R., Pelling, R. M., and Hurley, K. C. (1984), *Solar hard X-ray microflares*, Astrophys. J. **283**, pp. 421–425.

Lu, E. T. and Russell, R. J. (1991), *Avalanches and the distribution of solar flares*, Astrophys. J. **380**, pp. 89–92.

Maslov, S. and Sneppen, K. (2002a), *Specificity and stability in topology of protein networks*, Science **296**, art. no. 5569, pp. 910–913.

Maslov, S. and Sneppen, K. (2002b), *Topological pattern recognition in complex networks*, submitted to Phys. Rev. Lett.

Milgram, S. (1967), *The small world problem*, Psychol. Today **2**, pp. 60–67.

Newman, M. E. J. (2001), *Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality*, Phys. Rev. E **64**, art. no. 016132.

Newman, M. E. J. (2002a), *Assortative mixing in networks*, Phys. Rev. Lett. **89**, art. no. 208701.

Newman, M. E. J. (2002b), *The structure and function of networks*, Comput. Phys. Commun. **147**, pp. 40–45.

Newman, M. E. J. (2003a), *Mixing patterns in networks*, Phys. Rev. E **67**, art. no. 026126.

Newman, M. E. J. (2003b), *The structure and function of complex networks*, SIAM Review **45**, pp. 167–256.

Newman, M. E. J. (2004a), *Detecting community structure in networks*, Eur. Phys. J. B **38**, pp. 321–330.

Newman, M. E. J. (2004b), *Fast algorithm for detecting community structure in networks*, Phys. Rev. E **69**.

Newman, M. E. J. (2005), *Power laws, Pareto distributions and Zipf's law*, Contemp. Phys. **46**, pp. 323–351.

Newman, M. E. J. (2006), *Modularity and community structure in networks*, Proc. Natl. Acad. Sci. USA **103**, pp. 8577–8582.

Newman, M. E. J., Barabási, A.-L., and Watts, D. J., *The Structure and Dynamics of Networks* (Princeton University Press, 2006).

Newman, M. E. J., Forrest, S., and Balthrop, J. (2002), *Email networks and the spread of computer viruses*, Phys. Rev. E **66**, art. no. 035101.

Newman, M. E. J. and Girvan, M. (2004), *Finding and evaluating community structure in networks*, Phys. Rev. E **69**, art. no. 026113.

Pastor-Satorras, R. and Vespignani, A. (2001), *Epidemic spreading in scale-free networks*, Phys. Rev. Lett. **86**, pp. 3200–3203.

Pastor-Satorras, R., Vázquez, A., and Vespignani, A. (2001), *Dynamical and correlation properties of the Internet*, Phys. Rev. Lett. **87**, art. no. 258701.

Popper, K. R., *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge and Kegan Paul, London, 1963).

Price, D. J. (1965), *Networks of scientific papers*, Science **149**, pp. 510–515.

Price, D. J. (1976), *A general theory of bibliometric and other cumulative advantage processes*, J. Amer. Soc. Inform. Sci. **27**, pp. 292–306.

Scellato, S., Cardillo, A., Latora, V., and Porta, S. (2006), *The backbone of a city*, Eur. Phys. J. B **50**, pp. 221–225.

Shannon, C. E. and Weaver, W., *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

Simon, H. A. (1955), *On a class of skew distribution functions*, Biometrika **42**, pp. 425–440.

Smith, J. M. and Szathmáry, E., *The Major Transitions in Evolution* (W. H. Freeman/Spektrum, Oxford, 1995).

Snijders, T. A. B. (1991), *Enumeration and simulation methods for 0-1 matrices with given marginal*, Psychometrika **56**, pp. 397–417.

de Sola Pool, I. and Kochen, M. (1978), *Contacts and influence*, Soc. Netw. **1**, pp. 1–48.

Solomonoff, R. and Rapoport, A. (1951), *Connectivity of random nets*, B. Math. Biophys. **13**, pp. 107–117.

Takayasu, H. (1989), *Steady-state distribution of generalized aggregation system with injection*, Phys. Rev. Lett. **63**, art. no. 2563.

Trusina, A., Maslov, S., Minnhagen, P., and Sneppen, K. (2004), *Hierarchy measures in complex networks*, Phys. Rev. Lett. **92**, art. no. 178702.

Tyler, J. R., Wilkinson, D. M., and Huberman, B. A., *Email as spectroscopy: Automated discovery of community structure within organizations*, in *Proceedings of the International Conference on Communities and Technologies* (Kluwer Acadmics Publisher, 2003).

Vázquez, A., Pastor-Satorras, R., and Vespignani, A. (2002), *Large-scale topological and dynamical properties of the Internet*, Phys. Rev. E **65**, art. no. 066130.

Waldrop, M. M., *Complexity: The Emerging Science at the Edge of Order and Chaos* (Simon and Schuster, New York, 1992).

Watts, D. J. (2002), *A simple model of global cascades on random networks*, Proc. Natl. Acad. Sci. USA , pp. 5766–5771.

Watts, D. J., *Six Degrees: The Science of a Connected Age* (W. W. Norton and Co., New York, 2003).

Watts, D. J., Dodds, P. S., and Newman, M. E. J. (2002), *Identity and search in social networks*, Science **296**, pp. 1302–1305.

Watts, D. J. and Strogatz, S. H. (1998), *Collective dynamics of 'small-world' networks*, Nature **393**, pp. 440–442.

West, G. B., Brown, J. H., and Enquist, B. J. (1997), *A general model for the origin of allometric scaling laws in biology*, Science **276**, pp. 122–126.

Wilson, E. O., *Consilience: The Unity of Knowledge* (Alfred A. Knopf, 1998).

Wu, F., Huberman, B. A., Adamic, L. A., and Tyler, J. R. (2004), *Information flow in social groups*, Physica A **337**, pp. 327–335.

Zajdenweber, D. (1995), *Business interruption insurance a risky business. A study on some paretian phenomena*, Fractals **3**, art. no. 3, pp. 95–110.

Zipf, G., *Human behavior and the principles of least effort* (Addison Wesley, Cambridge, 1949).